2013

DISSECTION OF STRESS RESPONSE NETWORKS REGULATING MULTIPLE STRESSES IN RICE

Rafi Shaik
Michigan Technological University

DISSECTION OF STRESS RESPONSE NETWORKS REGULATING MULTIPLE STRESSES IN RICE

By

Rafi Shaik

A DISSERTATION

Submitted in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

In Biological Sciences

MICHIGAN TECHNOLOGICAL UNIVERSITY

2013

This dissertation has been approved in partial fulfillment of the requirements for the Degree of DOCTOR OF PHILOSOPHY in Biological Sciences

Department of Biological Sciences

Dissertation Advisor:	*Ramakrishna Wusirika*
Committee Member:	*Rupali Datta*
Committee Member:	*Guiliang Tang*
Committee Member:	*Hairong Wei*
Department Chair:	*Chandrasekhar P. Joshi*

Table of Contents

PREFACE

The material used in chapter 2 of this dissertation was reproduced from our article previously published in "*Plos One* **7**, e49331" under Creative Commons Attribution License (CCAL) open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. Entire data was produced in Dr. Wusirika laboratory in Department of Biological Sciences, Michigan Technological University. Dr. Wusirika was the corresponding author. He conceived, designed the experiments, contributed reagents/materials/analysis tools and wrote the paper. I, designed and executed the experiments, wrote and executed the computer programs and wrote the paper.

The data for the material used in chapters 3 and 4 of this dissertation was produced in Dr. Wusirika laboratory in the Department of Biological Sciences, Michigan Technological University. I conceived, wrote and executed computational programs and wrote the manuscript. Dr. Wusirika wrote and revised the manuscript.

ACKNOWLEDGEMENT

I would like to express my heartfelt gratitude to Dr. Ramakrishna Wusirika for accepting me to pursue my research career in his laboratory with his excellent guidance, caring and support towards my doctoral studies. It would not be possible to perform my experiments in the lab and publish the research in scientific journals without his timely advices, contribution and financial support.

I am sincerely grateful to my committee members Dr. Hairong Wei, Dr. Rupali Datta, and Dr. Guiliang Tang for their excellent guidance and support to finish my doctoral studies. I really appreciate their valuable time, suggestions and directions on my proposal, projects and dissertation.

I would like to thank Dr. Michael Gibson (Ex Chair), Dr. Chandrasekhar Joshi, Dr. Nancy Auer (graduate program director), Dr. Susan Bagley, Dr. Aparna Deshpande, Dr. Thomas Werner, Dr. Xiaoqing Tang, Dr. Michael Gretz, and all other faculty members from biological sciences.

I would like to express my gratitude to the staff members of Department of Biological Sciences: Jeff Lewin, Patricia Asselin and Emily Betterly and Department of Chemistry: Jerry Lutz and Andrew Galernaue for their indispensable assistance.

I am very much thankful to my friends, graduate students and postdoctoral fellows for their great help during my graduate studies. This list includes, but not limited to Ramana Reddy, Kefeng Li, Surendar Dhadi, Claire Doskey, Stephanie Groves, Aparupa Sengupta, Komal Kumar, Srinivas. R Sripathi, Emily Geiger, Giri Vegesna, Rajani Vegesna, Hansika Vegesana, Srinivas R. Mandalapu, Jithendar Rao Gujja, Jagadeesh Janjanam, Vimala Janjanam, and Navya Janjanam.

I would like to extend my gratitude to the graduate school dean, Dr. Jacqueline E. Huntoon, Dr. Debra, D. Charlesworth, and Nancy Byers Sprague for providing special fellowship to continue my graduate studies and excellent suggestions to finish my dissertation and degree.

Finally, I am forever indebted to my parents, family and relatives for their support to pursue my graduate studies at MTU.

ABSTRACT

Important food crops like rice are constantly exposed to various stresses that can have devastating effect on their survival and productivity. Being sessile, these highly evolved organisms have developed elaborate molecular machineries to sense a mixture of stress signals and elicit a precise response to minimize the damage. However, recent discoveries revealed that the interplay of these stress regulatory and signaling molecules is highly complex and remains largely unknown. In this work, we conducted large scale analysis of differential gene expression using advanced computational methods to dissect regulation of stress response which is at the heart of all molecular changes leading to the observed phenotypic susceptibility.

One of the most important stress conditions in terms of loss of productivity is drought. We performed genomic and proteomic analysis of epigenetic and miRNA mechanisms in regulation of drought responsive genes in rice and found subsets of genes with striking properties. Overexpressed genesets included higher number of epigenetic marks, miRNA targets and transcription factors which regulate drought tolerance. On the other hand, underexpressed genesets were poor in above features but were rich in number of metabolic genes with multiple co-expression partners contributing majorly towards drought resistance.

Identification and characterization of the patterns exhibited by differentially expressed genes hold key to uncover the synergistic and antagonistic components of the cross talk between stress response mechanisms. We performed meta-analysis on drought and bacterial stresses in rice and Arabidopsis, and identified hundreds of shared genes.

We found high level of conservation of gene expression between these stresses. Weighted co-expression network analysis detected two tight clusters of genes made up of master transcription factors and signaling genes showing strikingly opposite expression status.

To comprehensively identify the shared stress responsive genes between multiple abiotic and biotic stresses in rice, we performed meta-analyses of microarray studies from seven different abiotic and six biotic stresses separately and found more than thirteen hundred shared stress responsive genes. Various machine learning techniques utilizing these genes classified the stresses into two major classes' namely abiotic and biotic stresses and multiple classes of individual stresses with high accuracy and identified the top genes showing distinct patterns of expression. Functional enrichment and co-expression network analysis revealed the different roles of plant hormones, transcription factors in conserved and non-conserved genesets in regulation of stress response.

Chapter 1: Introduction and background

Plant stresses can dramatically alter plant growth, development and productivity. These stresses may not necessarily be immediately lethal, irreversible or occur permanently but depends on type, severity and duration of the stress. Any kind of plant stress factor can be broadly classified into one of the two following categories: abiotic stresses caused by non-living factors such as drought or flooding, intense sunlight or cold and biotic stresses caused by living organisms such as harmful insects or infectious bacteria resulting in significant deviation from optimal conditions for life of the plant (Vinebrooke et al., 2004). Early concepts of stress response or the effect of stress on plants was summarized as "state in which increasing demands made upon a plant lead to an initial destabilization of functions, followed by normalization and improved resistance, and if the limits of tolerance are exceeded and the adaptive capacity is overworked, the result may be permanent damage or even death" (Lichtenthaler, 1998) which still holds true although the latest findings differ in the complexity of stress response compared to what once was thought as limited and generic to a variety of stressors (Lichtenthaler, 1984).

Abiotic stress response are typically activated by stress signals such as hyperosmolarity caused by drought or salt stress that are perceived by sensors like histidine kinases (HKs) or Receptor Like Kinases (RLKs) or activation of ion channels on cell membranes that transduce the signal to intracellular compartments via MAPK cascades or Ca^{+2} signalling pathways (Chinnusamy et al., 2004) and activate many second messengers, plant hormones, signal transducers and transcriptional regulators. Two major kinds of abiotic transciptional regulatory networks are identified in drought, salt and cold stress namely, abscisic acid dependent pathways activated by stress-induced AP2 transcription factors (TFs) possessing the *cis*-acting element, ABRE (ABA-responsive element) which control the expression of MYB, MYC bZIP and NAC TFs, and ABA-independent pathways activated by TFs possessing the *cis*-acting element, DRE (dehydration-responsive element)/CRT (C-RepeaT) controlling the expression of HD-ZIP, AP2/ERF TFs (Shinozaki and Yamaguchi-Shinozaki, 2007). Expression of

these TFs leads to the synthesis of antioxidants, Late Embryogenesis Abundant (LEA) proteins, proteases and transporters that conferring osmotic homeostasis, damage repair and cellular protection (Hirayama and Shinozaki, 2010).

Biotic stress responses resulting from various kinds of plant-pathogen interactions may lead to compatible interaction where the host plant is unable to mount an effective anti-infectious defense response, allowing the pathogen to complete its life cycle or incompatible interaction where a series of complex defense responses are triggered activating a local response called Hypersensitive response (HR) or systematic long-term response. In HR, reactive oxygen species (ROS) levels quickly build up causing localized cell death. Systemic host responses, further classified as systemic acquired resistance (SAR) and induced systemic resistance (ISR) elevate levels of various phytohormones, protein kinases and antimicrobials which in-turn activate many downstream processes so that antimicrobial responses are activated more strongly in response to subsequent infection (Lodha and Basak, 2012).

Remarkable scientific breakthroughs revealing additional layers of regulation of gene expression like RNA interference, DNA methylation, histone modifications along with discovery of plethora of stress response factors including novel transcription factors, signaling molecules and small metabolites, and numerous ways they interact with each other provided a deeply intricate picture mounting to various forms of stress response known as stress escape or avoidance, tolerance and resistance (Hadiarto and Tran, 2011). For instance, submergence is one of the better understood stress conditions showing two antithetical adaptive responses, escape and tolerance, primarily governed by the multigenic *SNORKEL* (*SK*) and *SUBMERGENCE-1* (*SUB1*) loci (master regulators), respectively in rice, both of which encode tandem-repeated *ETHYLENE RESPONSIVE FACTOR* (*ERF*)-type transcription factor genes (Fukao and Xiong, 2013). While, SK induces gibberellic acid (GA)-mediated internode elongation enabling the plant to outgrow gradually rising floodwaters (submergence escape) (Hattori et al., 2009), SUB1 assists in endurance of complete submergence for weeks through restriction of carbohydrate consumption, chlorophyll degradation, and elongation growth

(submergence tolerance) (Xu et al., 2006). Evolution of these wide spectra of molecular programs equip the plants to sense change rapidly and adapt accordingly (Ahuja et al., 2010). The effectiveness of stress response varies widely depending on the species, genotype, tissue identity and developmental age of the plant. One of the efficient ways to dissect the effectiveness of stress response is to compare and uncover molecular basis of a physiologically stress tolerant plant variety against stress susceptible variety within the species (Huang and Guo, 2005, Sairam et al., 2005, Gorovits et al., 2007, Wang et al., 2007, Mizoi and Yamaguchi-Shinozaki, 2013, Narsai et al., 2013).

Losses in productivity of staple food crops due to biotic and abiotic stresses annually are estimated in the range of 30-60% in terms of potential yield or several billions of dollars in terms of economic return (Mittler and Blumwald, 2010, Seo et al., 2011). Further, global scarcity of water resources and the increased salinization of soil and water are posing serious threat to the food security of the world (Vinocur and Altman, 2005). By mid of 21st century, world population is expected to exceed 10 billion and witness serious shortage of food (Smith et al., 2010). Thus, development of crops that can sustain a wide range of stresses and still maintain high productivity is highly desired in order to meet various socio-economic and agro-economic challenges (Takeda and Matsuoka, 2008, Newton et al., 2011). Global efforts to achieve this goal are underway by both public and private enterprises. However, progress has been limited due to the highly complex nature of stress response where improvement of resistance to one stress resulted in reduction in its normal growth and development or productivity, for example, overexpression of transcription factors (TFs), AP59 and OsNAC6 resulted in improved drought tolerance but caused reduction in yield due to disruption in spikelet development (Oh et al., 2009) and showed dwarf phenotype, respectively (Hu et al., 2006, Nakashima et al., 2007). Further, enhancement of one stress was linked to increased susceptibility to other stresses, for instance, overexpression of stress-responsive Mitogen-activated protein kinase (MAPK) gene (*OsMAPK5*) from rice resulted in increased expression of pathogenesis-related (*PR*) genes such as *PR1* and *PR10* and significantly enhanced resistance to fungal (*Magnaporthe grisea*) and bacterial (*Burkholderia glumae*) pathogens but also caused significant reduction in drought, salt, and cold tolerance

(Xiong and Yang, 2003). Another study showed exogenous application of the phytohormone abscisic acid (ABA) enhanced low temperature tolerance in rice but made it more susceptible to its fungal pathogen *Magnaporthe grisea* (Koga et al., 2004).

The current understanding on the role of phytohormones in stress response is that ABA plays a central role in abiotic stress response and, salicylic acid (SA) and jasmonic acid (JA)/ethylene (ET) play major roles in biotic stress (Sharma et al., 2013). The interplay between the signaling pathways controlled by these phytohormones referred as signaling crosstalk largely contributes to the overlap observed between abiotic and biotic stresses suggesting the optimized use of molecular machinery to confront a wide range of stresses and various combinations of them which is poorly understood. The generation of reactive oxygen species (ROS) has been proposed as another key convergence point between biotic and abiotic stress responses (Fujita et al., 2006) due to their involvement in both ABA signaling and disease resistance responses (Gadjev et al., 2006). Based on the signals provided by ROS molecules and phytohormones, MAP-kinase cascades mediate the crosstalk. For instance in Arabidopsis, the cascade MEKK1–>MAPK kinase 2 (MKK2)–>MPK4/MPK6 was found to function as part of cold and salt stress signaling, while MEKK1–>MKK4/MKK5–>MPK3/MPK6 cascades have been reported to regulate the pathogen defense response pathway via the expression of WRKY22 and WRKY29. However, MPK3 and MPK6 were also found to be activated by abiotic stresses revealing their involved in both stress conditions (Droillard et al., 2002). The rice gene OsMPK5 is an ortholog of Arabidopsis MPK3 and was also reported to positively regulate tolerance of drought, salt, and cold stresses and negatively regulates pathogen resistance (PR) gene expression (Xiong and Yang, 2003).

Tremendous innovations in high throughput technologies and their widespread application to study different stresses and their combinations in model plants like Arabidopsis and food crops like rice, and free availability of raw data has propelled the field in to a fast track lane of -omics era, churning out massive amounts of bio-molecular information in the form of what can be called as stress –responsive epigenome, transciptome, proteome and metabolome identifying all the players, big and small and

inching us closer to the comprehensive understanding of the molecular architecture of stress response repertoire of plants. Bioinformatics methods and tools have become indispensible in storing, organizing and analyzing the deluge of multiple forms of bio-molecular data in recent years and in preparation for an information rich future. For instance, Gene Expression Omnibus is a central repository storing genome wide transcriptome data from microarray and sequencing experiments (Barrett and Edgar, 2006) along with ArrayExpress (Rustici et al., 2013) and Gene Expression Atlas (Kapushesky et al., 2012). Tools like GENEVESTIGATOR (Grennan, 2006), that provide a web-browser data mining interface to query these large microarray gene expression databases, PLAZA (Van Bel et al., 2012) to perform cross-species expression analysis, and PLANEX (Yim et al., 2013) to analyze co-expression networks are also available.

The imminent need to breed robust food and energy crops combined with emerging picture of complexity of stress responses and availability of multiple forms of high-throughput data provide impetus to systematically investigate the role of different regulatory layers and interaction among them to elicit a desired stress response. Towards this end, we analyzed the effect of different abiotic and biotic stress conditions on rice transcriptome. We chose rice because it is one of the most important staple food crops and a model plant species acting as a reference to a number of cereal and emerging biofuel grass species due to its 1) compact genome (~430Mb), 2) finished genome sequence of two subspecies (Goff et al., 2002, Yu et al., 2002), 3) extensive synteny and collinearity with other grass genomes (Feuillet and Keller, 2002), 4) availability of high density genetic maps and whole-genome microarrays (there are currently ~2000 microarray experiments done on just Affymetrix RiceArray chip available at GEO database), 5) well-established genetic transformation methods and availability of gene-indexed mutants for targeted loss-of-function or gain-of-function analysis of many rice genes (Jung et al., 2008) and, 6) multiple computational tools and databases developed to analyze rice specific bio-molecular data (Lee et al., 2009, Nagamura et al., 2011, Naika et al., 2013, Sato et al., 2013). RiceNet (Lee et al., 2011), for instance provides an

experimentally tested genome-scale gene network and was used to identify 13 novel genes involved in XA21 mediated immune response.

Comprehensive understanding of the regulatory networks involving molecular players from various layers such as epigenome, transciptome, proteome and metabolome that modulate the dynamic adaptive changes in a plant responding to stress is essential in developing robust food crops to meet the imminent energy demands of the future.

Chapter 2: Computational and proteomic analysis of epigenetic and microRNA mediated regulation of drought responsive genes in rice

Rafi Shaik and Wusirika Ramakrishna

The material contained in this chapter was previously published in "*Plos One* **7**, e49331"
http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0049331

2.1 Abstract

Drought stress response is a complex trait regulated at multiple levels. Changes in the epigenetic and miRNA regulatory landscape can dramatically alter the outcome of a stress response. However, little is known about the scope and extent of these regulatory factors on drought related cellular processes and functions. To this end, we selected a list of 5468 drought responsive genes (DRGs) of rice identified in multiple microarray studies and mapped the DNA methylation regions found in a genome wide methylcytosine immunoprecipitation and sequencing (mCIP-Seq) study to their genic and promoter regions, identified the chromatin remodeling genes and the genes that are targets of miRNAs.

We found statistically significant enrichment of DNA methylation reads and miRNA target sequences in DRGs compared to a random set of genes. About 75% of the DRGs annotated to be involved in chromatin remodeling were downregulated. We found one-third of the DRGs are targeted by two-third of all known/predicted miRNAs in rice which include many transcription factors targeted by more than five miRNAs. Clustering analysis of the DRGs with epigenetic and miRNA features revealed, upregulated cluster was enriched in drought tolerance mechanisms while the downregulated cluster was enriched in drought resistance mechanisms evident by their unique gene ontologies (GOs), protein protein interactions (PPIs), specific transcription factors, protein domains and metabolic pathways.

Further, we analyzed the proteome of two weeks old young rice plants treated with a global demethylating agent, 5-azacytidine (5-azaC), subjected to drought stress and identified 56 protein spots that are differentially expressed. Out of the 56 spots, 35 were differently expressed in the sample with both demethylation and drought stress treatments and 28 (50%) were part of DRGs considered in the computational analysis.

2.2 Introduction

In plants, epigenetic mechanisms including DNA methylation, histone modifications and certain small RNA (sRNA) mediated pathways regulate gene expression, chromatin structure and genome stability (He et al., 2011). Dynamic epigenetic changes in response to endogenous and external stimuli play a definitive role in the plasticity of phenotype of an organism adapting to adverse environmental conditions. Thus, an increasing number of studies with the aid of high-throughput sequencing and genome tilling microarray technologies are focusing on exploring the role of epigenetic mechanisms in genome evolution and ecological adaptation. A recent study revealed the global cytosine methylation patterns in rice using methylcytosine immunoprecipitation (mCIP) combined with Illumina sequencing (Yan et al., 2010). Genome-wide high resolution maps of DNase I hypersensitive (DH) sites from seedling and callus tissues of rice, which correlate with open chromatin structure revealed majority of DH sites to be located outside promoter regions and found 58% more DH sites in callus than in seedling (Zhang et al., 2012c). Small RNAs (sRNAs) are increasingly found to regulate the epigenome through chromatin based pathways for gene silencing (RNA directed DNA methylation pathway), paramutation, genetic imprinting and epigenetic reprogramming (Simon and Meyers, 2011). A study of S-locus protein 11 genes (*SP11*) of *Brassica* demonstrated that sRNA derived from the dominant *SP11* allele trigger methylation of the promoter of recessive *SP11* gene (Tarutani et al., 2010). While majority of sRNA in plants are small interfering RNAs (siRNAs) regulating transcriptional gene silencing, micro RNAs (miRNAs) play a key role in posttranscriptional gene silencing. Further, the distinction between siRNAs and miRNAs is becoming blurred, as both the molecules are intimately linked in terms of their origins and modes of operation (Voinnet, 2009). Thus, integration and analysis of data on differential gene expression, epigenetic and sRNA mediated regulation would reveal a comprehensive picture of the dynamics of stress responsive genome in generating phenotypic diversity and could have significant implications in agriculture.

Rice is one of the most important economically important cereal crops accounting for about one-fifth of the total caloric intake of the human population worldwide (Smith,

1995). Water deficit is a major abiotic factor affecting global crop yield and is known to induce a sequence of morphological, biochemical and molecular alterations that negatively affect plant growth and productivity (Wang et al., 2011a). With the advent of high-throughput technologies, dehydration tolerance in rice has been a subject of intense research resulting in a deluge of genomic, proteomic and metabolomic data (Choudhary et al., 2009, Ray et al., 2011, Shu et al., 2011b, Wang et al., 2011a). More than 5000 genes found to be differentially expressed in rice under drought stress by multiple studies were amalgamated by (Ray et al., 2011). Many of these drought responsive genes (DRGs) are either poorly annotated or very little is known about their regulatory control especially through epigenetic and miRNA mediated mechanisms. So far a few studies analyzed the role of epigenetic mechanisms in drought response in rice. A study between drought-tolerant and drought-sensitive rice lines found a difference of about 12% in genome wide DNA methylation/demethylation and they also reported 70% of these changes revert back to original status while 30% remain even after recovery (Wang et al., 2011b). Another study in rice has shown the differential expression of DNA methyltransferases in different developmental stages, tissues and abiotic stresses contributing to *de novo* DNA methylation and maintenance (Sharma et al., 2009). A genome wide miRNA study identified 30 miRNAs that are differentially expressed in drought response (Zhou et al., 2010).

In this study, we thematically collated and mapped the available information from different sources on DNA methylation; chromatin related proteins and sRNAs on DRGs and divided them into nine clusters based on presence/absence of these features and differential expression to pursue our goal of dissecting the orchestration of regulatory control in a plant cell responding to drought stress. Extensive characterization of the clusters based on a number of molecular features was performed. We also analyzed the proteome of young rice plants treated with 5-azacytidine (5-azaC) that causes global demethylation and grown in water deficit conditions to identify differentially expressed genes that are regulated by DNA methylation and play a role in drought response.

2.3 Methods

2.3.1 Drought responsive genes

The 5611 DRGs amalgamated by Ray et al. (Ray et al., 2011) from various drought studies on rice (Cooper et al., 2003, Rabbani et al., 2003, Zhou et al., 2007, Ray et al., 2011) were selected for this analysis. The 112 genes with affymetrix probe IDs mapping to *Oryza sativa* ssp *indica* were filtered out. The rest of the genes were matched with MSU release 7.0 of *Oryza sativa* ssp *japonica* (http://rice.plantbiology.msu.edu) and 31 obsolete loci from older MSU releases were also left out, leaving 5468 unambiguous DRGs belonging to *Oryza sativa* ssp *japonica* with latest annotation (Table S2.1). A random list of 5000 genes was generated using a Perl script from MSU7 annotation from all rice genes excluding pseudogenes and retrotransposon related genes. The 835 (15% of 5468) DRGs in the list were retained to truly account for randomness.

2.3.2 Epigenetic features

The mCIP-seq or DNA methylation reads in rice (Yan et al., 2010) were mapped on to the genomic location of the DRGs using a Perl script. The reads localized between transcriptional start site (TSS) and end of each gene with an overlap cut-off value of minimum 50 bases were collated and classified as genic DNA methylation reads and those falling 1kb upstream region of the TSS were collated and classified as promoter DNA methylation reads. The genes annotated as chromatin-associated proteins (CAPs) by the chromatin database (ChromDB) (Gendler et al., 2008) among the DRGs were identified. The plant miRNA database (PMRD) (Zhang et al., 2010) has 2641 miRNAs of rice (both experimental and computationally predicted miRNAs, including all the miRNAs reported in the miRBase database (Kozomara and Griffiths-Jones, 2011)) and their target genes predicted by psRNATarget server (Dai and Zhao, 2011). One or more micro RNAs (miRNAs) targeting each of DRGs as reported in plant microRNA database (PMRD) were identified.

2.3.3 Classification of DRGs into clusters

The DRGs were classified into nine clusters as shown in figure 2.1. All of the 5468 considered as part of the cluster D (Drought). The DRGs with any of the following features, DNA methylation reads overlapping promoter region or genic region, miRNA target and ChromDB gene were grouped together and classified as cluster E (epigenetic and miRNA) and those without any of the above features were classified as cluster NE (non epigenetic and miRNA). Each of the D, E and NE clusters were further classified into DU (drought upregulated) and DD (drought downregulated), EU (with epigenetic and miRNA features and upregulated) and ED (with epigenetic and miRNA features and downregulated), NEU (without epigenetic and miRNA features and upregulated) and NED (without epigenetic and miRNA features and downregulated) to reflect up or downregulation of gene expression.

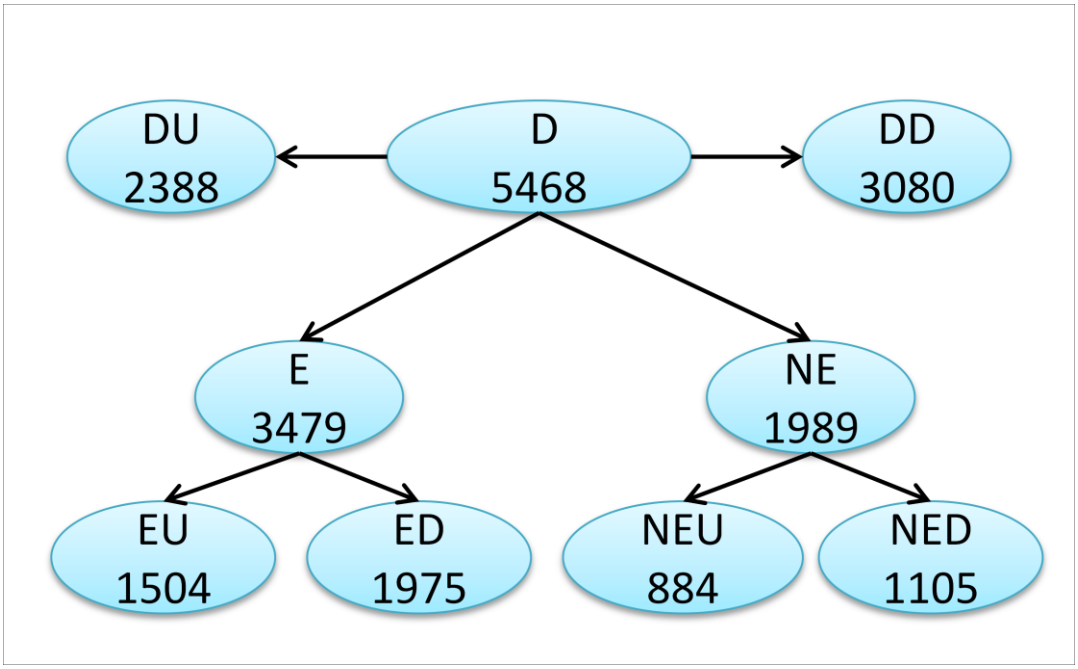

Figure 2.1: Classification of Drought Responsive Genes (DRGs) into nine clusters based on epigenetic/miRNA features and differential expression. Cluster D: all DRGs, DU: upregulated DRGs, DD: downregulated DRGs, E: Genes with any or all epigenetic/miRNA features, NE: no epigenetic/miRNA features, EU: E with upregulated DRGs only, ED: E with downregulated DRGs only, NEU: NE with upregulated DRGs only and NED: NE with downregulated DRGs only.

2. 3.4 Gene ontology analysis

The genes in each of the clusters were analyzed using the Singular Enrichment Analysis (SEA) tool by agriGO (Du et al., 2010) at default settings of Fisher t-test ($p<0.05$), False Discovery Rate (FDR) correction by Benjamini-Yekutieli method and five minimum number of mapping entries against species specific pre-computed background reference.

2. 3.5 Proteome analysis

The predicted protein-protein interactions (PPIs) shown by the protein(s) coded by every gene with all other protein(s) within the cluster were identified using the Search Tool for the Retrieval of Interacting Genes/Proteins database (String-DB) (Szklarczyk et al., 2011) with a combined score p-value<0.04.

The gene IDs annotated as members of different transcription factor (TF) families by plant transcription factor database (PlnTFDB v3.0, http://plntfdb.bio.uni-potsdam.de/v3.0/) (Perez-Rodriguez et al., 2010) were searched against the IDs of all DRGs. The plnTFDB had 3119 protein models belonging to 2399 genes annotated as TFs and were arranged in 80 families (TF families and other transcriptional regulators) for the species *Oryza sativa* subsp. *japonica*. Each of the TF family was analyzed to find the clusters they are enriched in. The lists of overlapping TF families in different clusters were analyzed using the tool Venny (Oliveros, 2007).

The protein domains present in all of the DRGs based on the classification by provided Pfam (Punta et al., 2012) were obtained from http://rice.plantbiology.msu.edu/ and were analyzed for overrepresented protein domains. Further, each of the domain family was analyzed to find the distribution of its members in the nine clusters.

The information about the metabolic pathway-associated genes was obtained from the data provided in RiceCyc (Jaiswal et al., 2006). Each pathway was analyzed to find the number of genes present in each of the clusters and the percentage of DRGs over total number of genes in that pathway.

2. 3.6 Drought stress and 5-azaC treatments

The protocol was adapted and modified from Boyko et al. (Boyko et al., 2010). The seeds of *Oryza sativa* ssp *japonica* obtained from the National Plant Germplasm System (NPGS) of the United States Department of Agriculture - Agricultural Research Service (USDA-ARS) were sterilized and germinated in a sterile petri plate wetted with half-Murashige and Skoog (MS) medium and grown in dark for 4 days at room temperature. Twenty young seedlings were transferred to magenta box each with 50ml of half-MS medium for control plants and 50ml of half-MS medium and 1-50 µM 5-azaC for treated plants (Thermo Fisher Scientific, NJ) and grown for two weeks in the dark at 28°C-day/25°C-night temperature, 12-h-light/12-h-dark cycle, and 50% humidity. Drought stress was given for 5hrs according to Dai et al. (Dai et al., 2007) by transferring the young plants to Whatman 3MM paper in a sterile petri dish.

2. 3.7 Two dimensional SDS-PAGE, in-gel digestion and MALDI-TOF

Total protein from four groups (control (C), drought stress (DS), 10µM 5-azaC (A) and 10µM 5-azaC with drought stress (ADS)) was isolated using ReadyPrep Protein Extraction Kit (Bio-Rad, CA) and quantified using BCA Assay. About 150µg of protein sample from each group was incubated in 200µl of rehydration buffer (8M urea, 2M thiourea, 2% CHAPS and 50mM DTT). Isoelectric focusing was carried out using 11cm immobiline dry strips (Bio-Rad, CA) with a non-linear pH 3-10 gradient. Strips were rehydrated using programmed voltage gradients at 20°C for a total of 12kVh and separated for 1h at 500V, 1h at 1000V, 2hrs at 6000V and 40min at 6000V. The IPG strips were reduced in equilibration buffer-I (0.375 M Tris-HCl, pH 8.8, 6M urea, 20% glycerol, 2% SDS, and 50mM DTT) for 20min at 25^{o}C and alkylated for 20min in equilibration buffer-II containing 150mM iodoacetamide. The equilibrated strips were placed on top of 15% polyacrylamide gels and run for 2.5hrs at 100V. Proteins were visualized by Coomassie Imperial Protein Stain (Pierce, Rockford, IL). Differentially expressed proteins between all groups were identified using ImageMaster (GE Healthcare Biosciences, PA).

Protein spots from 2D electrophoresis were excised from gels based on their fold change (>2-fold) and resolution. The gel pieces were destained twice with 200µl of 50% acetonitrile (MeCN)/25 mM NH_4HCO_3 buffer (pH 8.0) at room temperature for 20min, washed once with 200µl of 100% MeCN and vacuum dried by a SpeedVac concentrator (Savant, Holbrook, NY). The gel pieces were rehydrated with 13ng/µl sequencing grade modified trypsin (Promega; Madison, WI) in 25mM NH_4HCO_3 and incubated at 37°C overnight. Peptides were subsequently extracted twice with 50µl of 50% MeCN/5% formic acid for 15min at 37°C. All extracts were combined and dried. The peptides were eluted with 5µl of 75% MeCN/0.1% TFA. The peptides were analyzed using matrix assisted laser desorption/ionization time of-flight mass spectrometry (MALDI-TOF MS) (Microflex, Bruker). About 0.5 µl of 2,5-dihydroxybenzoic acid (DHB) matrix was loaded on a 96well ground steel MALDI plate followed by 0.5µl of peptide extract. Each sample was scanned with 1000 laser shots at 60% laser strength. The mass spectra were corrected for background subtraction and mass calibration. Protein masses were identified by searching NCBI_nr database through *MASCOT* search engine with 1 missed cleavage, ±100ppm of mass tolerance , carbamidomethylation of cysteines as fixed modification and oxidation of methionine as variable modification. *To identify the MSU7 IDs of the homologous proteins, BLASTP searches were performed (*http://rice.plantbiology.msu.edu/) and the best hits were selected.

2.4 Results

2.4.1 Epigenetic features of DRGs

A total of 2162 DRGs (39.5%) with one or more methylation reads (3633 total reads) falling in genic regions were identified (Fig. 2.2A), which is statistically significant (z-score: 2.58 at $p<0.05$) compared to a set of 5000 random genes. About 853 DRGs (40% of 2162) had more than one methylation read mapped to their genic region (Fig. 2.2B). The average gene length of the DRGs was 3522 bases while that of all genes in rice was 6656 (including transposon element (TE) genes). The average gene length of the DRGs with at least one methylation read in their genic regions was 4725 bases and those

without any methylation reads in their genic regions was 2735 bases. Our finding of significantly smaller average length of genes without any methylation reads (57% reduction) specifically among DRGs suggests a correlation between methylation and their gene size. Out of the 2162 DRGs, 461 (21.3%) had one or more methylation reads in the first 1000 bases from TSS. We identified 1249 (22%) and 913 genes (16.6%) with methylation reads in their genic region that were down and upregulated in drought stress, respectively. We found 678 DRGs (12.3%) with one or two mCIP-reads mapped to their promoter regions (Fig. 2.2A). Out of 678, 213 had methylation reads in the first 200 bases upstream of TSS. Interestingly, 296 (43%) DRGs with methylation reads mapped to their promoter region also had at least one methylation read mapped on to their genic region.

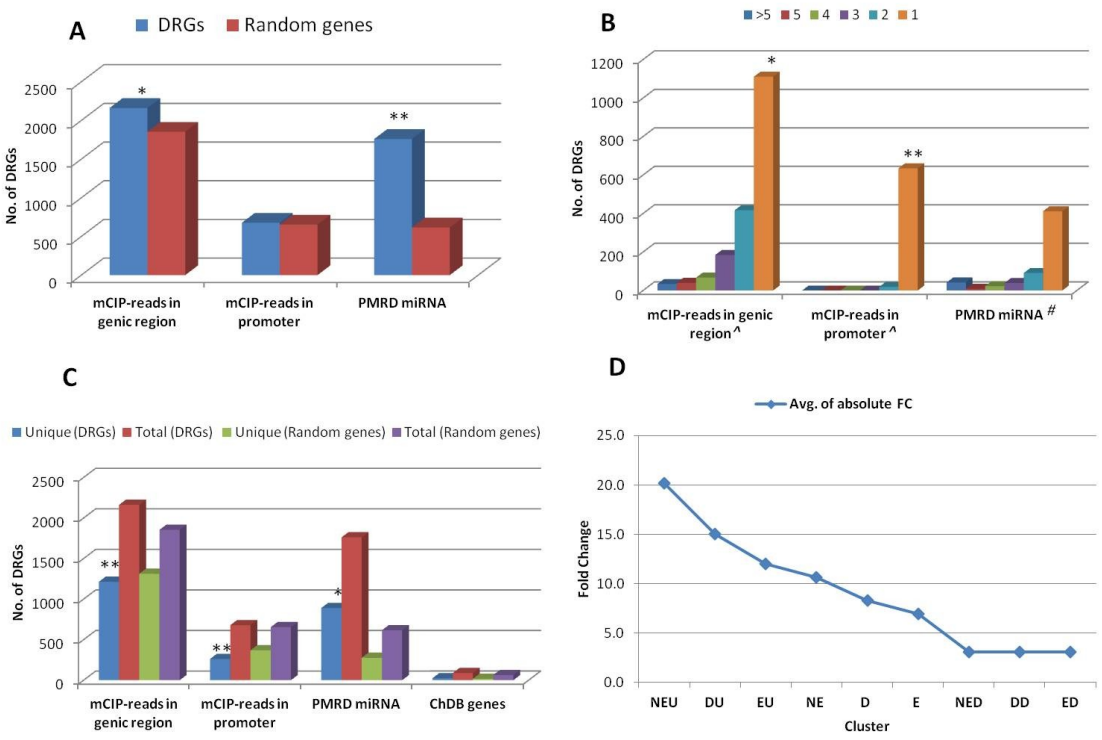

Figure 2.2: Epigenetic features of DRGs versus random set. A) The number of mCIP-reads mapped to the genic region of DRGs compared to the random set. In the same way mCIP-reads that mapped to promoter region were compared. Total numbers of miRNAs from PMRD database targeting the DRGs were compared to the random set. B) Distribution of multiple instances of epigenetic features on DRGs. ^Each of the bars represents number of DRGs mapped with given number of mCIP-reads only. #Each of the bars represents number of DRGs targeted by given number of miRNA only. C) Comparison of sets of genes with unique epigenetic features in DRGs with the random set. Unique represents the set of genes with only one of the three epigenetic/miRNA features and all represents number of genes with a particular feature and with one or more other epigenetic/miRNA features. D) Distribution of the average of absolute fold change of gene expression from (Ray et al., 2011) for the nine clusters. * indicates significant Z-score at $p<0.05$ and ** indicates significant Z-score at $p<0.01$.

In total, 1761 DRGs (32%) were potential targets of one or more miRNAs which is highly significant compared to the random set (616 or 12%) with a z-score of 24.25 ($p<0.01$) (Fig. 2.2A). A number of DRGs were predicted to be targets of multiple miRNAs (Fig. 2.2B). Ninety one DRGs were predicted to be targets of 10 or more miRNAs (Table S2.2). Three DRGs (LOC_Os08g13430 (expressed protein), LOC_Os05g18294 (SEC14 cytosolic factor family protein) and LOC_Os11g25780 (PB1 domain containing protein)) had more than 150 miRNAs targeting them. Out of 2641

miRNAs in PMRD, 1771 (67%) had at least one DRG as target with 82 miRNAs predicted to target 10 or more DRGs. The regulation of about one-third (32%) of DRGs by two-thirds (67%) of all known/predicted miRNAs reemphasizes the importance of miRNA mediated regulation of these DRGs and the need to comprehensively understand their mechanism of action. The miRNAs, osa-miRf10273-akr predicted by miMatcher pipeline (Lindow et al., 2007) and osa-miR414 experimentally identified in the moss *Physcomitrella patens* (Fattash et al., 2007), were predicted to target highest number of DRGs (103 and 75, respectively). We found 88 DRGs (17% of 514 rice genes in ChromDB) that are chromatin related genes. Interestingly, 66 of these 88 DRGs (75%) were downregulated suggesting that the chromatin landscape of the rice genome has been dramatically altered in drought response.

The DRGs with only one of the three epigenetic features studied were analyzed and compared to a random set of 5000 genes (Fig. 2.2C). The DRGs with DNA methylation in either genic or promoter region seem more likely to share other epigenetic features. This is evident by the significant negative z-score of -9.5 and -6.4 ($p<0.01$) for the number of DRGs with only DNA methylation in genic region and only DNA methylation in promoter region, respectively as the epigenetic feature compared to the random set. The number of DRGs targeted by miRNA exclusive of other epigenetic features is 890 (16% of all DRGs) while that for random set is 276 (~5%) (z-score 2.45 $p<0.05$). The number of ChromDB genes exclusive of other features is 25 while that of random set is 18. The number of DRGs with DNA methylation in genic region which are also targets of miRNA are 736 (13% of all DRGs) while the same for random set is 299 (5%) (z-score 12.8 at $p<0.01$). Similarly, number of DRGs with DNA methylation in promoter region which are also targets of miRNA are 219 (4%) while the same for control set is 63 (1.2%) (z-score 8.6 at $p<0.01$). The number of DRGs having DNA methylation in genic and promoter regions and also are targets of miRNA (PMRD) (all three epigenetic features) are 104 (1.9%) while the number in random set is 26 (0.5%) (z-score 6.3 at $p<0.01$).

2.4.2 Cluster analysis of DRGs elucidates different gene expression patterns

Overall, cluster E had 63.6% of all DRGs and the clusters with genes that are downregulated had higher number of genes even upon classifying into sub clusters (Clusters DD, ED, NED compared to DU, EU and NEU) (Fig. 2.1). Comparison of average of the absolute fold change of gene expression of each of the clusters showed a clear trend of higher fold change for all the clusters with upregulated genes (EU, DU and NEU) and lower fold change for all the clusters with downregulated genes (ED, DD and NED) (Fig. 2.2D). The positioning of the cluster NEU at top as shown in figure 2D, suggests that the genes in cluster NEU could be expressed through a simpler route as they are not under direct control of epigenetic and miRNA mediated mechanisms. On the other hand, the lowest average fold change of gene expression of cluster ED could possibly be due to tighter control of the genes in this cluster and are very selectively expressed, specifically in stress response.

2.4.3 GO analyses of the clusters reveal a number of novel biological processes and functions of DRGs

In total, we found 1011 significant GOs ($p<0.05$) for all of the nine clusters combined. These comprised 320 unique GOs out of which 189, 90 and 41 were related to biological process (BP), molecular function (MF) and cellular component (CC), respectively. Out of these 73 GOs (22.9%) are unique to only one of the 9 clusters (Table S2.3). Besides reporting most of the GOs that are known to be enriched in DRGs by other studies, our analysis revealed a vast number of novel GOs as a result of clustering based on the underlying regulatory information. For instance, the GO "response to biotic stimulus" was found to be significant ($p=0.00026$) only in cluster D. Even upon classifying the cluster D into clusters DD and DU this term was not significant. Conversely, the GO "ncRNA metabolic process" was found to be significant ($p=0.0021$) only in cluster ED and was not significant in other clusters including cluster D. A few more examples showing enrichment/depletion of GOs in DRGs due to clustering are illustrated in figure 4.

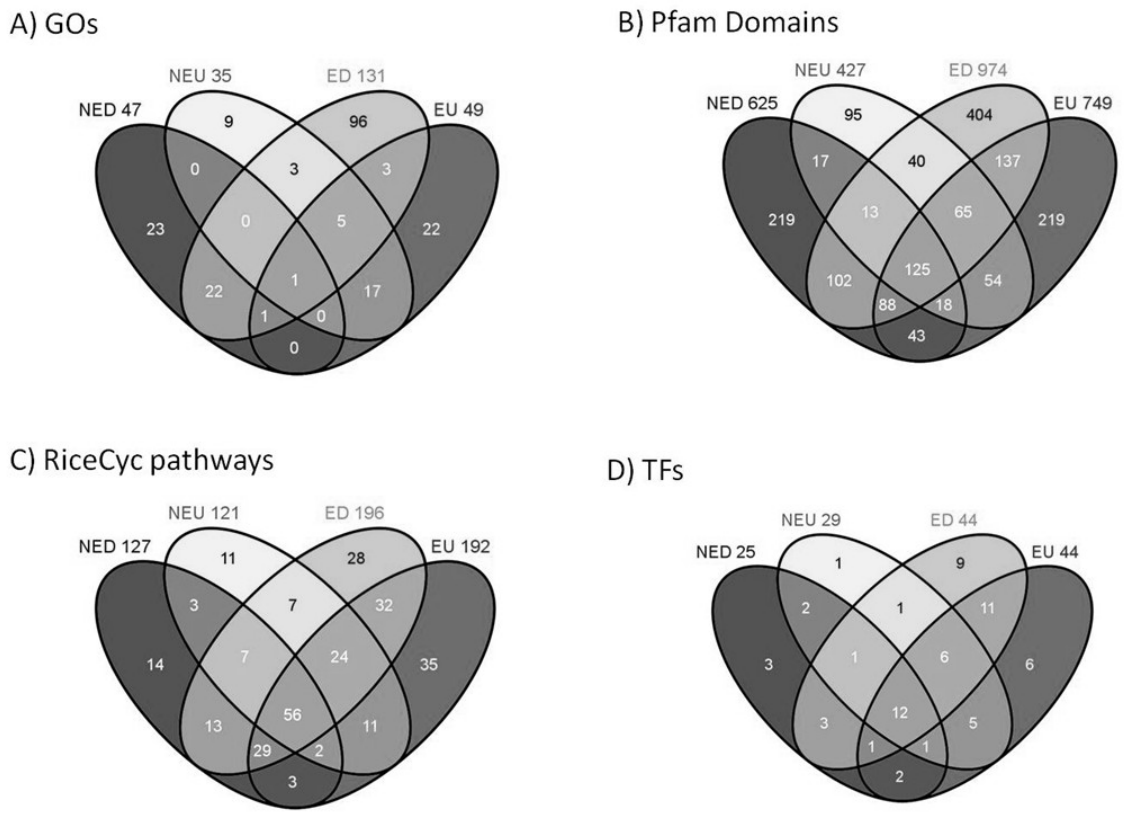

Figure 2.3: Four way venn diagrams depicting overlap of different characteristics between the clusters NED, NEU, ED and EU. A) GO terms analyzed by AgriGO, B) Protein domain families as per Pfam database, C) Metabolic pathways by RiceCyc and D) Types of transcription factors as reported by PlnTFDB.

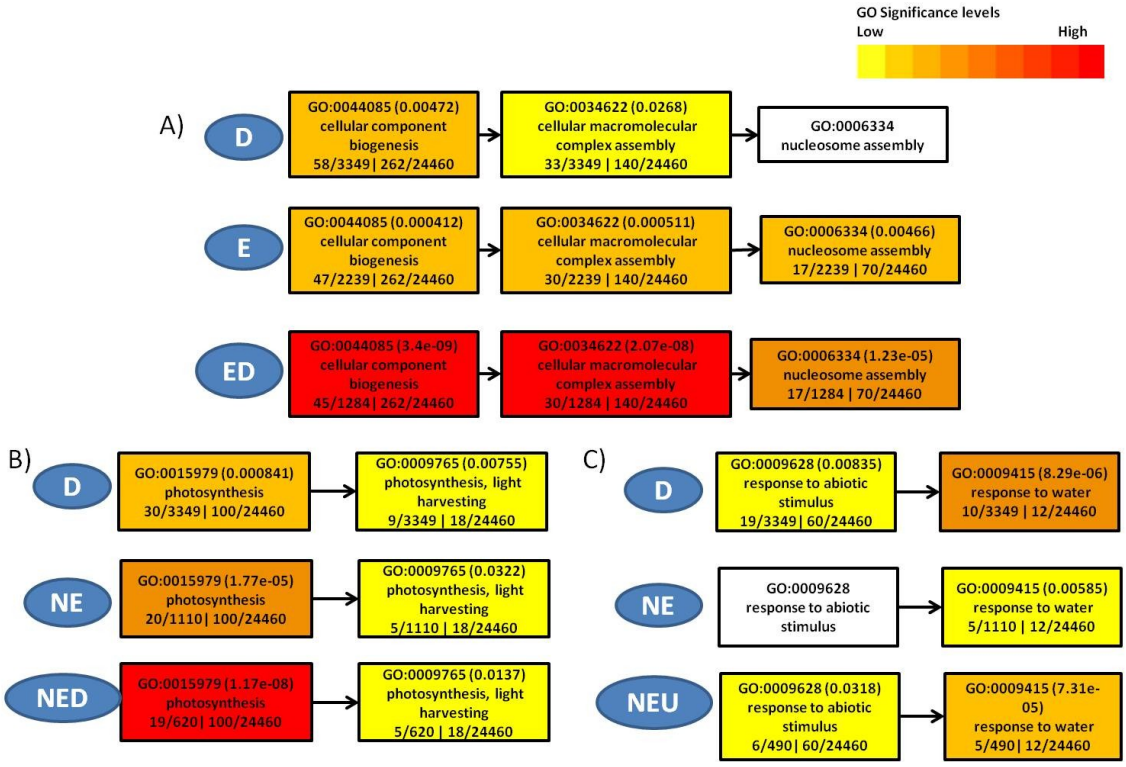

Figure 2.4: Few examples showing changes (increase/decrease) in significance of GO terms as we move down from a large cluster to sub clusters. Also shown are the changes in significance of GO terms as we move from a parent to child GO term (indicated by the direction of the arrows). A) and B) show increase and C) shows decrease in the significance of GO terms as we move to sub clusters.

Each of the clusters revealed distinct GOs that clearly define their properties. Significant overlapping GOs were observed in the groups that are either up or downregulated such as 22 common and exclusive GOs between NED and ED and 17 between NEU and EU (Fig. 2.3A and Table S2.4). On the other hand, there were no shared GOs common and exclusive between NED and NEU and only 3 GOs between ED and EU. About 73% of GOs of the ED are unique to ED (overrepresented) and a major portion of the remaining GOs were shared with cluster NED exclusively. A number of GOs that are unique to NED are related to photosynthesis such as "photosystem", "photosynthetic membrane", "photosynthesis light harvesting" and other terms include "structural molecule activity", "protein folding", and "response to oxidative stress". Conservation of energy by reduction of photosynthetic activity and translation are known drought response

mechanisms. Exclusive enrichment of these processes in the cluster NED suggests they are probably not under direct epigenetic and miRNA control.

Cluster NEU shows a peculiar behavior of not overlapping with NED with no common GOs in 3 out of 4 possible combinations which suggests the clear demarcation of processes controlled by genes that belong to NEU and NED (Fig. 2.3A). Out of the 14 common GO BP terms between clusters NEU and EU, 11 are related to regulatory processes (Table S2.4).The GOs "response to water" ($p<0.00009$) and "response to abiotic stimulus" ($p<0.0007$) were also common to NEU and EU. The GOs unique to cluster NEU are mostly related to "RNA biosynthesis", "metabolism", "transcription" and "regulation of these processes" (Fig. S2.1). This result is in agreement with the expectation that genes involved in processes like RNA biosynthesis and transcription perform basic housekeeping functions of the cell and do not require subtle control by higher order regulatory mechanisms. Yet, upregulation of genes with these functions suggest requirement of the cell under stress to produce a large quantities of different kinds of RNAs as part of drought response.

The GOs that are unique to the EU overall seem to be related to protein modification processes especially "serine/threonine phosphatase activity" which is enriched significantly ($p=8.00E-08$) in addition to "signal transduction processes" and "response to osmotic stress". Reduction of transpiration by stomatal closure and accumulation of osmoprotectants in response to the resulting osmotic stress are well known mechanisms of drought response. Cluster ED with highest number of significant GOs is also the cluster with highest number of non-overlapping GOs (96/131 GO terms or 73%). This cluster shows a high number of terms related to nucleosome and cytoskeletal reorganization, and metabolic processes implying the complex regulation of energy conservation mechanisms by downregulation of a number of metabolic processes and reorganization of a number of cellular structures inside the cell responding to drought.

Table 2.1: List of consensus co-expression modules found in each stress geneset

MSU ID	PPI Count	Gene Description	DNA methylation and miRNA features
Cluster ED			
LOC_Os07g32880	84	ATP synthase gamma chain, putative, expressed	chr7_solexa13_1006351 (p)
LOC_Os10g10180	80	methyltransferase domain containing protein, putative, expressed	osa0miRf118720akr
LOC_Os01g03040	80	expressed protein	chr1_solexa12_1000315; chr1_solexa12_1000316; (g)
LOC_Os04g41340	78	4-nitrophenylphosphatase, putative, expressed	osa0miRf108630akr
LOC_Os08g06530	75	rubredoxin family protein, putative, expressed	chr8_solexa13_1000861 (g)
LOC_Os12g38640	75	expressed protein	chr12_solexa13_1007345 (g)
LOC_Os07g07540	74	SHOOT1 protein, putative, expressed	osa0miRf102730akr; osa0miRf109470akr
LOC_Os08g07060	73	CRR6, putative, expressed	chr8_solexa13_1000916 (g); osa0miRf115530akr
LOC_Os02g01150	73	erythronate-4-phosphate dehydrogenase domain containing protein, expressed	osa0miRf118380akr
LOC_Os02g47020	71	phosphoribulokinase/Uridine kinase family protein, expressed	chr2_solexa13_1008132; chr2_solexa13_1008133 (g)
Cluster EU			
LOC_Os01g14440	28	WRKY1, expressed	chr1_solexa12_1002043 (g)
LOC_Os05g46760	26	STE_MEKK_ste11_MAP3K.19 - STE kinases include homologs to sterile 7, sterile 11 and sterile 20 from yeast, expressed	chr5_solexa13_1007768 (g); osa-miRf12002-akr
LOC_Os05g25920	18	expressed protein	chr5_solexa13_1004521 (p); osa-miRf10947-akr
LOC_Os03g17700	18	CGMC_MAPKCGMC_2_ERK.2 - CGMC includes CDA, MAPK, GSK3, and CLKC kinases, expressed	chr3_solexa12_1002251 (g)
LOC_Os08g38210	18	transcription factor BIM2, putative, expressed	chr8_solexa13_1007354 (g)
LOC_Os04g52840	18	tyrosine protein kinase domain containing protein, putative, expressed	chr4_solexa13_1009408 (g)
LOC_Os06g44250	17	haemolysin-III, putative, expressed	osa-miRf12029-akr
LOC_Os01g61080	17	WRKY24, expressed	osa-miRf10947-akr
LOC_Os10g42690	16	jmjC domain containing	JMJ706 (ChromDB ID);

		protein, expressed	osa-miRf10002-akr
LOC_Os02g13840	16	citrate synthase, putative, expressed	chr2_solexa13_1001684 (p)
Cluster NED			
LOC_Os04g51792	72	PAP fibrillin family domain containing protein, expressed	N.A
LOC_Os02g42570	69	ferredoxin-thioredoxin reductase, variable chain, putative, expressed	N.A
LOC_Os01g68450	67	expressed protein	N.A
LOC_Os03g17070	63	ATP synthase B chain, chloroplast precursor, putative, expressed	N.A
LOC_Os03g16050	62	fructose-1,6-bisphosphatase, putative, expressed	N.A
LOC_Os10g15300	60	expressed protein	N.A
LOC_Os08g27010	59	APE1, putative, expressed	N.A
LOC_Os01g55570	58	expressed protein	N.A
LOC_Os02g51820	57	expressed protein	N.A
LOC_Os07g13969	55	expressed protein	N.A
Cluster NEU			
LOC_Os01g64470	13	harpin-induced protein 1 domain containing protein, expressed	N.A
LOC_Os01g72530	12	OsCML31 - Calmodulin-related calcium sensor protein, expressed	N.A
LOC_Os06g04240	12	expressed protein	N.A
LOC_Os06g10210	11	expressed protein	N.A
LOC_Os03g53020	9	helix-loop-helix DNA-binding domain containing protein, expressed	N.A
LOC_Os10g25290	9	ZIM domain containing protein, putative, expressed	N.A
LOC_Os06g46950	8	EF hand family protein, putative, expressed	N.A
LOC_Os11g10470	8	expressed protein	N.A
LOC_Os04g43680	7	MYB family transcription factor, putative, expressed	N.A
LOC_Os03g60570	7	ZOS3-22 - C2H2 zinc finger protein, expressed	N.A

* g and p in brackets denote that the methylation read(s) overlap genic and promoter regions respectively.

2.4.4 Clusters EU, ED, NEU and NED exhibit distinct patterns of protein-protein interactions

The STRING database analyses revealed higher number of interactions in downregulated clusters with 4.69 and 5.86 PPIs per gene with at least one PPI in NED and ED, respectively, compared to the upregulated clusters with 1.37 and 1.85 PPIs per gene with at least one PPI in NEU and EU, respectively (Fig. 2.5). About 35% of genes show PPIs in clusters NED and ED while only about 14.5% and 19.6% of genes show PPIs in clusters NEU and EU, respectively. This suggests the probable location of a number of down regulated genes at the bottom of regulatory cascades as evident by their significant GOs related to multitude of processes including metabolic, biosynthetic, and photosynthesis processes which involve many kinds of PPIs to synthesize or degrade a number of metabolites/biological substances and upregulated genes at the top of regulatory cascades controlling a few critical reactions as supported by the fact that 11 out of 14 GO BPs common and exclusive to EU and NEU are related to regulatory processes and they also show high average of the absolute fold change of gene expression. A list of top ten DRGs with highest number of PPIs in the clusters EU, ED, NEU and NED are given in table 2.1. The top ten DRGs of EU contain three major TF, three major kinase family genes and a jmjC domain coding gene which regulates chromatin reorganisation processes (Klose et al., 2006) suggesting that these genes mediate some of the most important drought response reactions. The complete PPI network of EU is shown in Fig. 2.6 and the individual PPIs along with their String-DB score is given in table S2.5. . Out of the 295 DRGs in EU, 115 had only one PPI and 14 had >6 PPIs (Fig. 2.6B). Two DRGs, LOC_Os01g14440 (OsWRKY1v2 - superfamily of TFs having WRKY and zinc finger domains) and LOC_Os05g46760 (STE_MEKK_ste11_MAP3K.19- STE kinase, part of the MAPK signaling cascade) had 28 and 26 PPIs each with other members in EU. Both the genes have one DNA methylation read overlapping with their genic regions and LOC_Os05g46760 is also predicted as a target of osa-miRf12002-akr.

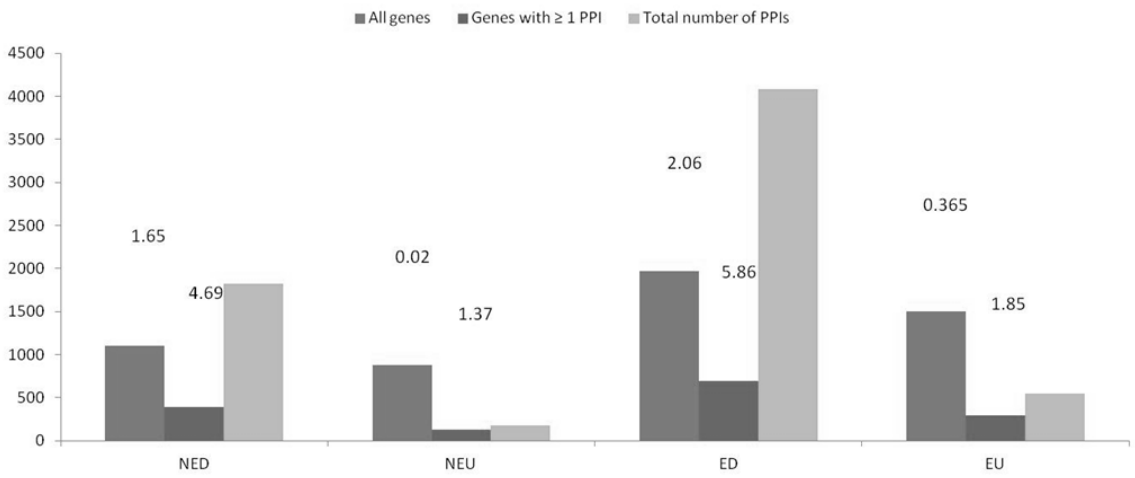

Figure 2.5: Number of protein protein interactions (PPIs) found in the four leaf clusters. The numbers above the bars represent average number of PPIs per gene over total number of PPIs found in the cluster and average number of PPIs per gene among genes with ≥1 PPI over total number of PPIs found in the cluster.

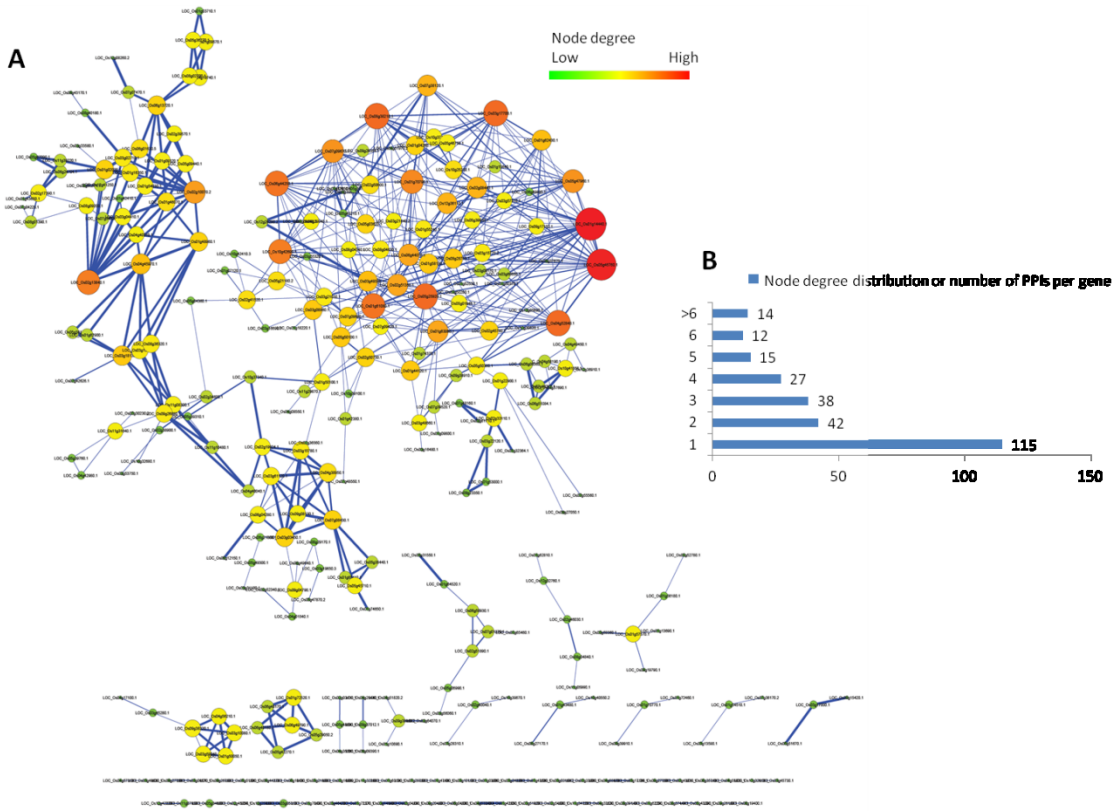

Figure 2.6: Protein-protein interactions (PPIs) network of cluster-EU. A) Network diagram showing DRGs as circles with size and color corresponding to number of PPIs. Higher the no. of PPIs larger the circle and thicker the color. Thickness of edges or connections between two nodes is based on String-DB score. Higher the String-DB score, thicker the connection. B) The number of genes with different number of PPIs (X-axis)

2.4.5 Characterizing the DRG clusters based on transcription factor family distribution

Out of the 5468 DRGs, 450 (8%) were annotated as transcription factor genes (Table 2.2). Interestingly, these 450 Drought Responsive Transcription Factors (DRTFs) represent most of the TF families (64/80 families in PlnTFDB) (Table S2.6). Although the cluster size of DU was smaller than DD, higher numbers of DRTFs were present in DU. Similarly NEU had the highest percent of TFs even though it had the least number of genes among all the nine clusters and NED had the least percent of TF genes. These results are similar to the trends observed in our fold change analysis (Fig. 2.2D).

Table 2.2: Number of transcription factor genes in the nine clusters

Cluster	No. of genes	No. of TF genes	Percent of TF genes in the group
Rice genome	55986	2399	4.28
D	5468	450	8.23
NE	1989	148	7.44
E	3479	302	8.68
NED	1105	45	4.07
NEU	884	103	11.65
ED	1975	144	7.29
EU	1504	158	10.51
DD	3080	190	6.17
DU	2388	261	10.93

Majority of the members of important TF families AP2-EREBP (29 out of 38 DRTF genes), bHLH (20/32), bZIP (19/27), C3H (9/9), DBP (3/3), HSF (9/10), LOB (5/6), NAC (22/30), PHD (6/7), Tify (6/7) and Trihelix (5/5) were in cluster DU while majority of the members of TF families CCAAT (7/9), G2-like (9/11) and MADS (14/18) were in ED (Table S2.6). The number of TF families that are unique and common between different leaf clusters is shown in figure 3D. AP2-EREBP is one of the largest TF families unique to plants and is characterized by the presence of AP2 DNA-binding domain. AP2-EREBP has the highest number of DRTFs and majority of the members (75%) are upregulated in drought response suggesting upregulation of a number of functional roles attributed to this family. Similar trend was shown by other large TF families, namely NAC and bZIP. A number of major TF families were exclusively found in EU and ED (MADS, C2C2-CO-like, CCAAT, and HMG). Many TF families show bias to one of the four clusters. For example, 12/21 WRKY DRTFs are present in EU and 14/18 MADS DRTFs are present in ED. One of the major role played by MADS box genes is development of plant reproductive structures, specifically floral meristem and organ identity (Riechmann and Meyerowitz, 1997). The enrichment of MADS TFs in ED suggests that these mechanisms are subtly controlled and downregulated as part of drought resistance to conserve energy.

Table 2.3: Number of genes with pFAM domains and the number of different domain families found in the nine clusters

Cluster	No. of genes	No. of genes with one or more Pfam domains	No. of domain families
Rice genome	55986	33779	3337
D	5468	4348	1639
NE	1989	1477	879
E	3479	2871	1308
NED	1105	829	625
NEU	884	648	427
ED	1975	1617	974
EU	1504	1254	749
DD	3238	2446	1271
DU	2388	1902	915

2.4.6 Protein domain family distribution analysis reveals enrichment of major domain families in clusters with epigenetic and miRNA features

Rice genome has 33779 genes with at least one Pfam domain belonging to 3337 families. Out of 5468 DRGs, 4348 have Pfam domains belonging to 1639 families (Table 2.3) suggesting the wide range of changes involved in drought response. Overall the clusters E and DU show significantly higher percentage of genes with at least one Pfam domain compared to NE and DD (cluster-E 82.5%, G-statistic-51.4 and DU 79.6%, G-statistic-13.1 compared to NE 74.2% and DD 75.5% respectively). Figure 3B shows the number of domain families that are unique and common across different combinations of clusters. The trends observed here are similar to those in figure 3A with NED and ED, NEU and EU, and ED and EU showing higher overlap than other cluster combinations.

A number of major domain families were enriched in cluster E suggesting many proteins with functional roles in signal transduction and metabolism are under direct epigenetic control. For example 182/1144 Pkinase domains in rice are found in DRGs out of which 138 or 75% were present in cluster E (Table S2.7). The other domains showing similar trend in cluster E include LRR_1 (63/87 domains found in all DRGs), NB_ARC (23/32), SRF-TF (18/18), peptidase_S10 (14/15) and terpene_synth (11/13). Further, all of the above mentioned domains were significantly enriched in cluster ED suggesting the processes controlled by these domains are highly downregulated in drought response.

Many other domains that were enriched in cluster E were further enriched in cluster EU. Examples include zf-C3HC4 (40/63 domains in all of DRGs were part of E out of which 26 were part of EU), PP2C (20/29 E and 17/29 EU) and raffinose_syn (5/5 E and 4/5 EU). Raffinose family oligosaccharides (RFO) were found to act as ROS scavengers and also play a role in protection against freezing, desiccation and high temperature stress (Bolouri-Moghaddam et al., 2010). All seven dehydrin domain containing genes found in the rice are part of the DRGs considered in this study and all of the seven were upregulated. Dehydrin domain containing proteins are produced in plants in response to low temperatures and drought stress and protect membranes from damage (Puhakainen et al., 2004).

A number of DUFs (Domain of Unknown Function) also showed enrichment in distinct clusters, suggesting that these domains could be playing an important role in drought response that is unknown. For example, 8/11 DUF221 domains were part of DRGs out of which 7 were part of DU and 5 were part of EU. The only annotation available for DUF221 is that it is a family of hypothetical transmembrane proteins (http://Pfam.sanger.ac.uk/family/PF02714). A number of domains and families although present in high numbers in rice, were found to be underrepresented in DRGs including zf-CCHC, hATC, chromo, Peptidase_C48 and FAR1 domains (Table S2.7).

2.4.7 Metabolic pathway analysis reveals enrichment of pathways involved in synthesis of a number of amino acids, peptides and sugars in cluster EU which function as osmoprotectants and antoxidants

We found 275 out of 357 pathways listed in RiceCyc to be differentially regulated in drought stress (Table S2.8). The distribution of the pathways in different DRG clusters is shown in table S2.9. About 20% of 275 pathways were common to all of the four leaf clusters NED, NEU, ED and EU (Fig. 2.3C). Approximately 35% of the pathways are exclusive to cluster E while only 10% are unique to NE. DRGs involved in amino acid synthesis pathways including proline, alanine, citrulline, methionine were significantly enriched in cluster EU (Table S2.8), which are known to serve as osmoprotectants and antioxidants as part of drought response. (Kawasaki et al., 2000, Akashi et al., 2001).

Glutathione (GSH) is a tripeptide known to act as a redox sensor for environmental stress. Antioxidant defense reactions, which use GSH as an electron donor for the regeneration of ascorbate are considered as the main pathway of superoxide and H_2O_2 removal (Kumar et al., 2010) . We found gamma-glutamyl cycle and ascorbate biosynthesis pathways to be enriched in the cluster EU. Trehalose functions as a stress protectant, stabilizing proteins and membranes against destruction (Garg et al., 2002). Multiple genetic studies have proposed trehalose pathway as a central metabolic regulator (López-Gómez and Lluch, 2012). There are 19 genes linked to trehalose biosynthesis I pathway in RiceCyc, out of which 7 are part of the DRGs. We found all of the 7 DRGs to be part of DU and 5 to be part of EU. The DRGs encoded for the enzymes involved in GDP-D-rhamnose and GDP-L-fucose synthesis which are components of primary cell wall were also found to be significantly enriched in EU. Jasmonic acid is a hormone known to induce lipoxygenases that protect against membrane alterations during water stress (Rock et al., 2010) . Twelve out of 13 DRGs found to be involved in jasmonic acid biosynthesis pathway were under epigenetic control and eight of those were enriched in cluster EU.

A number of biosynthetic pathways were found to be specifically enriched in ED including those related to fatty acids, nucleotides, sugars like sucrose and UDP-D-xylose, cellulose, heme, lysine, phenylpropanoid and folate derivatives. While the degradation pathways of amino acids like tryptophan and valine were enriched in ED, their biosynthetic pathways were enriched in EU. All eight of the DRGs involved in tRNA charging pathway were part of DD and 7 of which were also part of ED. Out of the 26 genes involved in photorespiration, 10 are DRGs and 7 of which were part of ED. A number of basic metabolic pathways were significantly enriched in cluster E but dispersed between the clusters EU and ED including biosysnthesis and degradation pathways of glucose, galactose and starch, TCA cycle, biosynthesis of phospholipids, lipoxygenases (LOX), brassinosteroids, cysteine, methionine and degradation pathway of sucrose.

2.4.8 Proteome analysis of 5-azaC treated and drought stressed rice identifies epigenetically regulated DRGs

To identify proteins whose corresponding genes are regulated by DNA methylation and play a role in drought stress, rice seedlings were treated with 5-azaC and subjected to drought stress. Varying concentration of 5-azaC were tested and 10μM 5-azaC was selected because concentrations of >20μM drastically reduced the growth of rice seedlings. Two-dimensional gel electrophoresis analysis of total protein extract identified 201, 411, 205 and 501 differentially expressed spots with a fold change value of ≥2 in the control (C), control with drought stress (DS), 5-azaC (A), and 5-azaC and drought stress (ADS) samples, respectively when compared with each of the other three samples. Out of these, we analyzed 75 spots chosen based on both high fold change value and resolution for precise spot elution from the gel and successfully identified 56 proteins (Table 2.4). *T*here are multiple transcription factors, kinase/phosphatases, signaling, metabolic, structural proteins and also nine proteins annotated as 'expressed proteins' and six proteins related to retrotransposons in the identified protein list. Except eight spots which were differentially expressed between samples C and DS, the other 48 spots (86%) were differentially expressed in relation to samples treated with 5-azaC. GO analysis of these 48 spots identified 9 genes to be involved in stress response and 5 genes in protein modification processes (Fig. S2.2). We identified 35 proteins that are differentially expressed (25 upregulated and 10 downregulated) in sample ADS when compared against the other three samples. Comparison of ADS against DS revealed 11 upregulated and 3 downregulated proteins (Table 2.4).

Out of the 56 identified proteins, 28 (50%) were part of DRGs considered in our cluster analysis. Nine out of the 25 proteins upregulated in the sample ADS were part of cluster ED, 5 out of which were upregulated in comparison to the sample DS. Among the five genes, the gene coding for lactate/malate dehydrogenase had a mCIP-read mapped to its promoter and the other four coding for pentatricopeptide repeat (PPR) domain containing protein, calreticulin precursor protein, sucrose-phosphate synthase and glutathione S-transferase, respectively were targets of one or more miRNAs. We also found the genes coding for DnaK (Hsp70) family protein and laccase precursor protein which are part of

cluster EU to be overexpressed in ADS. The above findings suggest that these genes known to be up or downregulated under drought stress were upregulated in ADS due to the deregulation of their own methylation state or genes regulating them.

Similarly, out of the 10 genes upregulated in sample A, five were part of DRGs in cluster analysis. Out of the five genes, four were part of cluster DD and were upregulated compared to in DS or ADS samples. LOC_Os01g48874 (cluster ED) coding for wax synthase was upregulated in sample A compared to the sample DS. LOC_Os01g48874 had a mCIP-read mapped to its promoter suggesting probable activation of this gene due to demethylation. Out of the 9 spots upregulated in sample C, 3 were part of DRGs and all three were upregulated in comparison to sample DS. While the gene LOC_Os03g40830 was part of cluster ED, LOC_Os08g38210 and LOC_Os08g39840 coding for transcription factor BIM2 and LOX9, respectively were part of cluster EU.

2.5 Discussion

Our workflow pipeline of integrating genome wide epigenetic and miRNA data over DRGs, clustering and characterizing the subsets of genes with different types of molecular features revealed a number of novel insights about the key stress responsive regulatory modules. One of our major objectives was analyzing DRGs that are under epigenetic/miRNA control as clusters and contrasting them against DRGs that are not under epigenetic/miRNA control. This resulted in revelation of a comprehensive list of molecular mechanisms and pathways (tables S2.4-S2.8) that are specific (unique to highly enriched) to the genes that are under epigenetic/miRNA control. This information resource significantly enhances our know-how of drought stress regulation. Our other objectives include providing a readily searchable database of DRGs with epigenetic and miRNA data, provided in table S2.1. Identify key DRGs based on connectivity with other DRGs and functional significance within sub clusters (Fig. 2.6A and Tables 2.1, S2.5). Identify drought responsive proteins (DRPs) that are regulated by DNA methylation and compare them with DRGs clusters (Table 2.3).

Statistically significant enrichment of features like DNA methylation reads in genic region, miRNA target sequences in DRGs compared to a random set of genes suggests

DRGs are under tight epigenetic control. The negative z-score of DRGs with only DNA methylation reads in promoter or genic regions compared to random set (Fig. 2.2C) suggests co-occurrence of these regulatory features with other epigenetic features and can act as one of the metrics to determine the significance a DRG based on how tightly it is regulated. In our analysis we found a number of subsets of DRGs showing striking enrichment of certain features. For instance, 75% of the DRGs annotated to be involved in chromatin remodeling were downregulated. This set of genes can be further explored in determining the fitness of a drought responsive phenotype. Another interesting set of genes are the 1761 DRGs which are 32% of all DRGs considered in this study but targeted by 67% of all known/predicted miRNAs in rice which include many transcription factors targeted by more than five miRNAs, while the random set had only 12% of genes that are miRNA targets.

We found a number of DRGs with meager annotation that might be playing an important role in drought response. There are 989 DRGs (18% of all DRGs) with gene description as 'expressed protein' or 'hypothetical protein'. Out of these, 806 genes (15% of all DRGs) do not have any GO annotation mappings. This fact reveals that there is still a lot that is not known about drought response in Rice. In our analysis, presence of these genes in sub clusters that are associated with specific biological processes provides clues about their functional role. Epigenetic/miRNA features of these DRGs provide ways to manipulate their gene expression which could aid in determining their functions and also possibly identify new drought related mechanisms. For example LOC_Os03g15033 is annotated as an expressed protein with domain DUF3353. This gene is downregulated in drought stress (cluster ED) and is targeted by the highest number of miRNA (20 miRNAs that are part of miRBase). Our results reveal the key control switches and global scale regulatory dynamics that can be potentially engineered to further enhance the process of drought adaptation for genes that are well characterized including some that have shown improvement in transgenic drought adaption. Table 2.5 shows a list of ten DRGs belonging to cluster EU, on which transgenic studies were conducted. These DRGs showed improvement in drought tolerance by transgenic upregulation of ABA-dependent signaling transduction pathway, dehydrin family proteins, LEA proteins, seed

storage/lipid transfer proteins, transcription factors, protein kinases, cell membrane stability-related proteins and phosphatases, increased grain yield, polyamines and osmolyte synthesis, decreased cuticular permeability and reduced water loss. Overexpression of two genes (LOC_Os11g03370 and LOC_Os01g66120) which are part of the cluster NEU code for NAC transcription factors, showed improvement in drought tolerance in transgenic studies (Hu et al., 2008, Zheng et al., 2009).

Different molecular features that we analyzed for the leaf clusters are summarized in Table 2.6. Overall the cluster EU seems to be made up of DRGs that mediate drought tolerance mechanisms involving osmotic adjustments, antioxidant activities, desiccation tolerance etc (Hadiarto and Tran, 2011) evident by the seven late embryogenesis abundant (LEA) genes, GOs that are unique to the cluster including protein modification and signal transduction processes (Table S2.4), high average fold change of gene expression (Fig. 2.2D), high number of TFs (Table 2.2), less number of PPIs (Fig. 2.5), enrichment of protein domains including PP2C, zf-C3H4, raffinose_syn, methyltransf_29 (Table S2.7) and pathways related to synthesis of amino acids, peptides and sugars which are osmoprotectants, antioxidants, protein and membrane stabilizers (Table S2.8). On the other hand the cluster ED seems to be made up of DRGs that mediate processes related to drought resistance involving earliness to drought response, reduced leaf area, leaf rolling, reduced tillering, stomatal closure, efficient roots, reduced transpiration, etc (Hadiarto and Tran, 2011) evident by highest number of unique GO terms (73%) including nucleosome and cytoskeletal organization, majority of metabolic processes, lowest average fold change of gene expression, low number of TFs but significant enrichment of MADS-box TFs that control flowering genes among others, high number of PPIs, enrichment of p450, helicase, LRR_1 domains and enrichment of a number of biosynthesis pathways resulting in cellular adjustments and energy conservation.

We performed 2D-PAGE analysis of rice seedlings subjected to partial demethylation and drought stress to test the overall effect of epigenetic mechanisms on DRGs, specifically to analyze if there is a reversal in the differential expression of the DRGs as a result of demethylation of the promoter or gene sequence. Among the 28 proteins that

matched to the DRGs of our cluster analysis there are 15 ED and 7 EU genes. Eight out of the 15 genes in cluster ED have methylation sequences in their genic or promoter regions and are overexpressed in samples compared to those subjected to drought stress. The reversal in the expression of these genes is likely due to demethylation effect of 5-azaC. LOC_Os10g33800 coding for lactate/malate dehydrogenase has a methylation read mapped to its promoter and is highly overexpressed (fold change 38.94) in sample ADS compared to DS. Similarly, LOC_Os01g48874 coding for wax synthase has four methylation reads mapped to its genic region and is highly overexpressed (fold change 22.85) in sample A compared to DS. The differential expression of many other genes in sample ADS which do not possess methylation reads in their promoter or genic regions indicates the possibility of their regulated by other genes whose methylation state was altered due to the 5-azaC treatment. Thus, careful analysis of the identified genes would reveal the extent of role of epigenetic regulation in drought stress response.

Although many of the DRGs are extensively annotated and our analysis revealed key regulatory switches for the DRGs based on current status quo on epigenetic and miRNA mediated regulation, we expect comprehensive annotation (including siRNA, chromatin modifications and possibly other mechanisms yet to be discovered) of all the DRGs would enrich or deplete some of the striking patterns found in the clusters based on different molecular features. Thus, our study represents a first step towards the understanding of global regulatory control of stress response through integration of multiple annotation resources and unraveling a number of subsets of genes involved in key regulatory modules which could be further explored.

2.6 **Conclusion**

Our analysis of DRGs as clusters based on epigenetic and miRNA features dissected biological processes and molecular functions that play a key role in the regulation of stress response. We found a number of subsets of genes showing significant enrichment of certain characteristics suggesting that these set of genes can be further studied to explore their role as regulatory modules in drought response. Understanding the influence of these regulatory modules on transcriptional/post-transcriptional gene

silencing/activation and long term stress memory would be critical in engineering a drought sensitive plant variety with desirable traits into a drought resistant variety.

Table 2.4. Differentially expressed protein spots found in 5-azaC and drought treated samples

Spot I.D	MSUv7 ID	MSU Gene Product Name	Fold change*				Cluster	Coverage	Mascot score
			C	DS	A	ADS			
C-17	LOC_Os03g38840	retrotransposon, putative, centromere-specific			2.86		-	40.20%	43.9
C-19	LOC_Os04g16830	DNA-directed RNA polymerase subunit beta, putative			9		-	18.70%	53.9
C-212	LOC_Os07g22720	2-oxo acid dehydrogenases acyltransferase domain containing protein		4.2			-	13.20%	84.2
C-220	LOC_Os08g38210	transcription factor BIM2, putative		3.14			EU	10.40%	64
C-241	LOC_Os08g19680	expressed protein		3.3			-	52.90%	45.6
C-260	LOC_Os03g40830	OsSub30 - Putative Subtilisin homologue, expressed		4.8			ED	6.00%	71
C-266	LOC_Os08g39840	lipoxygenase, chloroplast precursor, putative, expressed		3.91			EU	43.90%	48.3
C-615	LOC_Os10g35412	retrotransposon protein, putative, unclassified, expressed				3.03	-	10.70%	58.3
C-99	LOC_Os09g32670	UDP-glucuronate 4-epimerase, putative		4.25			-	21.00%	58.6
DS-107	LOC_Os10g21190	expressed protein				4.9	-	73.00%	60.5
DS-109	LOC_Os11g15570	Ser/Thr protein phosphatase family protein, putative			4.55		NED	11.20%	64
DS-14	LOC_Os03g39010	possible lysine decarboxylase domain containing protein, expressed			34.72		-	25.40%	68.6
DS-187	LOC_Os11g10480	dehydrogenase, putative				5.15	EU	13.70%	62.5
DS-19	LOC_Os03g10510	outer mitochondrial membrane porin, putative				2.41	ED	18.60%	66.7
DS-206	LOC_Os01g07120	AP2 domain containing protein, expressed	3.26				EU	14.90%	51.4
DS-278	LOC_Os03g07700	expressed protein	3.05				NEU	10.40%	65
DS-32	LOC_Os08g41620	ubiquitin carboxyl-terminal hydrolase family protein, expressed			5.37		-	36.50%	72
DS-36	LOC_Os09g25270	hypothetical protein			5.62		-	22.30%	71.5
DS-41	LOC_Os12g37400	MCM7 - Putative minichromosome maintenance MCM complex subunit 7			5.6		NED	41.30%	51
DS-64	LOC_Os12g39830	cyclin, putative				5.05	ED	23.60%	107

DS-81	LOC_Os02g41800	auxin response factor, putative		5.55		-	16.30%	99.3
A-108	LOC_Os08g39150	expressed protein	4.55			EU	14.50%	70
A-133	LOC_Os01g31220	expressed protein			5.91	-	14.20%	77.9
A-164	LOC_Os03g62290	expressed protein	6.36			-	30.90%	66
A-21	LOC_Os01g48874	wax synthase, putative	22.85			ED	18.70%	65
A-23	LOC_Os03g06540	retrotransposon protein, putative, unclassified, expressed	6.33			-	18.40%	62.2
A-234	LOC_Os09g04440	DNA-binding protein, putative			6.33	ED	13.20%	67
A-516	LOC_Os04g38810	formin, putative, expressed			2.04	ED	6.20%	71
A-676	LOC_Os12g10670	AAA-type ATPase family protein, putative			4.8	NED	26.20%	42.8
A-730	LOC_Os12g13780	retrotransposon protein, putative, Ty1-copia subclass, expressed			3.27	-	41.30%	56.3
A-93	LOC_Os11g09070	expressed protein	2.5			-	17.80%	60.6
ADS-144	LOC_Os04g44224	brain acid soluble protein 1, putative	3.93			-	29.40%	65
ADS-188	LOC_Os09g37670	expressed protein	2.21			-	56.00%	53.9
ADS-198	LOC_Os10g33800	lactate/malate dehydrogenase, putative	38.94			ED	28.30%	77.9
ADS-20	LOC_Os01g36600	PPR repeat domain containing protein, putative	6.86			ED	76.20%	47.9
ADS-212	LOC_Os04g40950	glyceraldehyde-3-phosphate dehydrogenase, putative	2.37			-	33.50%	131
ADS-292	LOC_Os07g14270	calreticulin precursor protein, putative	2.25			ED	33.50%	123
ADS-373	LOC_Os11g47760	DnaK family protein, putative	5.26			EU	13.70%	131
ADS-393	LOC_Os01g66730	exosome complex exonuclease RRP40, putative		9.51		-	23.40%	82
ADS-484	LOC_Os02g45950	expressed protein	30.82			-	15.70%	87
ADS-503	LOC_Os10g21268	ribulose bisphosphate carboxylase large chain precursor, putative	7.87	25.21		-	36.60%	179
ADS-546	LOC_Os03g08170	protein kinase APK1B, chloroplast precursor, putative	2.2			ED	70.90%	64.2
ADS-549	LOC_Os04g18660	retrotransposon protein, putative, Ty3-gypsy subclass, expressed	5.26			-	9.30%	69.4
ADS-574	LOC_Os03g16610	laccase precursor protein, putative	5.21			EU	9.80%	64
ADS-578	LOC_Os12g44170	pentatricopeptide, putative	2.02			-	6.90%	86

ADS-687	LOC_Os09g38710	HEAT repeat family protein, putative			3.37	-	42.50%	56.3
ADS-695	LOC_Os05g44720	retrotransposon protein, putative, unclassified, expressed			3.31	-	55.90%	61.9
ADS-699	LOC_Os11g47970	AAA-type ATPase family protein, putative	3.98			ED	41.40%	125
ADS-701	LOC_Os07g47230	TKL_IRAK_DUF26-lh.10 - DUF26 kinases have homology to DUF26 containing loci	16.4	12.55		-	9.00%	70.6
ADS-712	LOC_Os04g52000	protein phosphatase 2C, putative			4.51	NEU	18.10%	74
ADS-725	LOC_Os09g39180	RNA recognition motif containing protein, putative			2.27	NED	23.90%	76.5
ADS-74	LOC_Os12g19381	ribulose bisphosphate carboxylase small chain, chloroplast precursor, putative			3.45	ED	84.40%	71
ADS-742	LOC_Os08g08060	vacuolar protein sorting-associated protein 18, putative			5.05	-	5.30%	80
ADS-748	LOC_Os01g54080	kinesin motor protein-related, putative			3.11	ED	10.90%	75.9
ADS-81	LOC_Os01g69030	sucrose-phosphate synthase, putative		2.42		ED	8.30%	90.7
ADS-87	LOC_Os10g38580	glutathione S-transferase, putative		24.45		ED	22.20%	77

*- The fold change (overexpression) value of each spot in column Spot I.D compared to the samples in columns C, DS, A and ADS.

Table 2.5. DRGs in cluster EU that showed improvement in drought tolerance in transgenic studies

Gene	Common name	Gene description	Epigenetic/miRNA features*	Reference
LOC_Os06g10880	OsABF2	bZIP transcription factor	chr6_solexa13_1001253 (g)	(Hossain et al., 2010)
LOC_Os02g08230	OsGL1-2	WAX2	chr2_solexa13_1000885 (g)	(Islam et al., 2009)
LOC_Os02g50350	OsDHODH1	dihydroorotate dihydrogenase protein	osa-miRf10310-akr	(Liu et al., 2009)
LOC_Os11g29870	OsWRKY72	WRKY72	chr11_solexa14_1005447 (g); osa-miRf10273-akr;osa-miRf10576-akr	(Yu et al., 2010)
LOC_Os06g04070	OsAdc1	pyridoxal-dependent decarboxylase protein	chr6_solexa13_1000396 (p); osa-miR1848; osa-miR815a; osa-miR815b; osa-miR815c	(Capell et al., 2004)
LOC_Os02g12310	OsDREB1A	no apical meristem protein	chr2_solexa13_1001389 (g)	(Ito et al., 2006)
LOC_Os01g58420	AP37	AP2 domain containing protein	chr1_solexa12_1009761 (p)	(Oh et al., 2009)
LOC_Os02g43970	ARAG1	AP2 domain containing protein	chr2_solexa13_1007628 (p)	(Zhao et al., 2010)
LOC_Os02g52780	OsbZIP23	bZIP transcription factor	chr2_solexa13_1008914 (g)	(Xiang et al., 2008)
LOC_Os05g46480	OsLEA3-1	late embryogenesis abundant protein, group 3	osa-miRf11013-akr	(Xiao et al., 2007)

Table 2.6. Comparision of different molecular features found in the leaf clusters EU, ED, NEU and NED

	EU	ED	NEU	NED
Average of absolute fold change	12	3.06	20.16	3.09
mCIP-reads in promoter region*	280 (18.6%)	398 (20%)	0	0
mCIP-reads in genic region	913 (60%)	1249 (63%)	0	0
PMRD miRNA targets	771 (51%)	990 (50%)	0	0
miRBase miRNA targets	163 (10.8%)	229 (11.5%)	0	0
ChromDB annotated genes	22 (25%)	66 (75%)	0	0
Unique GO terms among leaf clusters ^	22 (48%)	96 (73%)	9 (25.7%)	23 (49%)
Genes with PPIs within the cluster (String-DB)	296 (19.6%)	697 (35%)	129 (14.5%)	389 (35%)
TF genes (PlnTFDB)	158 (10.5%)	144 (7.2%)	103 (11.6%)	45 (4%)
Pfam domain containing genes	1254 (63.4%)	1617 (81.8%)	648 (73%)	829 (75%)

Metabolic pathways (RiceCyc) unique to the cluster among the leaf clusters	35 (18%)	28 (14%)	11 (9%)	14 (11%)
Genes found in 5-azaC drought study among the identified protein spots §	7 (12%)	15 (25.8%)	2 (3%)	4 (6.8%)

* percentage is no. of genes with the feature over total no. of genes in the cluster;
^ percentage is over total no. of GO terms found in the cluster; § percentage over total identified protein spots

Chapter 3: Genes and Co-expression Modules Common to Drought and Bacterial Stress Responses in *Arabidopsis* and Rice

Rafi Shaik and Wusirika Ramakrishna

The material contained in this chapter is under review with the PLoSOne

3.1 Abstract

Plants are simultaneously exposed to multiple stresses resulting in enormous changes in the molecular landscape within the cell. Identification and characterization of the synergistic and antagonistic components of stress response mechanisms contributing to the cross talk between stresses is of high priority to explore and enhance multiple stress response. To this end, we performed meta-analysis of drought (abiotic), bacterial (biotic) stress response in rice and *Arabidopsis* by analyzing a total of 386 microarray samples belonging to 20 microarray studies and identified approximately 3100 and 900 DEGs in rice and *Arabidopsis*, respectively. About 38.5% (1214) and 28.7% (272) DEGs were common to drought and bacterial stresses in rice and Arabidopsis, respectively, majority of which showed conserved expression status in both stresses. Gene ontology enrichment analysis clearly demarcated the response and regulation of various plant hormones and related biological processes. Fatty acid metabolism and biosynthesis of alkaloids were upregulated and, nitrogen metabolism and photosynthesis was downregulated in both stress conditions. WRKY transcription family genes were highly enriched in all upregulated gene sets while 'CO-like' TF family showed inverse relationship of expression between drought and bacterial stresses. Weighted gene co-expression network analysis divided DEG sets into multiple modules that show high co-expression and identified stress specific hub genes with high connectivity. Detection of consensus modules based on DEGs common to drought and bacterial stress revealed 9 and 4 modules in rice and *Arabidopsis* respectively with conserved and reversed co-expression patterns.

3.2 Introduction

Crop productivity and survival is tightly linked to its environment which is being altered due to climate change, biodiversity loss and degradation of land and freshwater (Foley et al., 2011) threatening the food security of the world while the food demand is estimated to increase by 70% in 2050 (Tester and Langridge, 2010, Godfray, 2011, Reynolds et al., 2012). According to latest World Agricultural Supply and Demand Estimates (WASDE) report by United States Department of Agriculture (USDA), about 80% of agricultural land is experiencing drought and over 2,000 U.S. counties had been designated as disaster areas (WASDE, 2012). Reflecting the declining environmental conditions, more often than not plants today are exposed simultaneously to multiple stresses resulting in enormous changes in the molecular landscape within the cell. Comprehensive understanding of the regulatory networks that modulate the dynamic adaptive changes in a plant responding to stress is critical to meet future energy needs. Rice and *Arabidopsis* are both model plant organisms representing monocots and dicots respectively. Both the plants have extensive biological knowledgebase and resources including complete genome sequence and highest number of microarray studies in the plant kingdom. Thus, analysis of stress responsive genes within and between rice and *Arabidopsis* for different kinds of stresses would reveal a number of pivotal attributes spanning across the major plant division, angiosperms.

Advancements in high throughput technologies have resulted in deluge of various kinds of -omic data addressing different aspects of temporal and spatial response in variety of stresses in plants. Microarray technology revolutionized the identification of global transcriptomic changes and today multiple transcriptomic studies exist for the same or related stress conditions. Thus meta-analysis of related microarray studies is increasingly becoming popular to enhance the sensitivity of the hypothesis addressed and validate conclusions (Tseng et al., 2012). So far, very few meta-analysis studies are available in plant systems (Adie et al., 2007, Ghanekar et al., 2008, Meier et al., 2008, Cohen et al., 2010, Finka et al., 2011). Meta-analysis of microarray data from *Arabidopsis* infected with eight different viruses revealed hub genes that are highly connected, organized in modules and are central to plant defense response (Rodrigo et al.,

2012). It is reported that in plants responding to multiple stresses, there exists extensive cross-talk between different stress responses via hormonal signaling pathways (Seki et al., 2002). Thus, it is imperative to compare and analyze different kinds of stress responses to find the genes, proteins and metabolites that are common and specific to different kinds of abiotic and biotic stress conditions. Meta-analysis of microarray studies involving samples from a wide range of tissues, developmental stages and different levels of stresses but specific to one stress condition would unravel the universal principles and features related to the stress response. Comparative analysis of such universal molecular profiles from different stresses would allow the identification of unique and shared features. Further, comparison of the stress responsive profiles across diverse plant species would reveal the conserved stress specific mechanisms and uncover orthologous genes that are most critical to the stress response.

Recently, there has been an upsurge in the number of studies reporting global co-expression networks of plants based on genome wide transcriptome data (Ficklin et al., 2010, Mochida et al., 2011, Downs et al., 2013). A number of tools namely ATTED-II (Obayashi et al., 2009), CressExpress (Srinivasasainagendra et al., 2008), RiceArrayNet (Lee et al., 2009), OryzaExpress (Hamada et al., 2011) and RiceFREND (Sato et al., 2013) based on co-expression networks are available that can be explored to identify novel genes, predict gene functions and characterize gene regulatory networks. A network based analysis in rice identified drought responsive gene modules and found a module with 134 genes specifically associated with both drought tolerant and drought resistant rice varieties (Zhang et al., 2012b). Weighted Gene Co-expression Network Analysis (WGCNA) is one of the latest and popular methodologies to decipher correlation patterns across microarray samples (Langfelder and Horvath, 2008). Implemented in R as a package, WGCNA provides a vast array of functions to detect, analyze and export individual and consensus modules from diverse but related microarray studies. WGCNA has been utilized to detect coexpression modules in *Arabidopsis*, rice, maize, soybean and poplar (Childs et al., 2011, Weston et al., 2011, Downs et al., 2013) and also across species (Ficklin and Feltus, 2011).

In this study, we performed large scale comparative transcriptomic analysis via meta-analysis of microarray data on drought and bacterial stress in rice and *Arabidopsis*. To elucidate the cross talk between different stress conditions, knowledge of the expression status of genes involved in stress response is critical. Our analysis revealed the genes that are unique to each stress and those that are shared with other stress conditions. Further, within common genes, we also found genes that were up or downregulated in both stresses and also genes which showed reversed expression status. Extensive analysis of various gene sets based on Gene Ontologies (GO), KEGG Orthologies (KO) and metabolic pathway analysis unraveled the underlying biological mechanisms related to different stresses. We then performed co-expression network analysis which divided the stress responsive genes into tightly co-expressed modules revealing organization of stress transcriptome.

3.3 Methods

3.3.1 Selection of stress related microarray studies

Gene Expression Omnibus (GEO) is the central repository for microarray and other forms of high-throughput data (Barrett et al., 2007). Experiments conducted on the Affymetrix platforms, Rice Genome Array (GPL2025) and *Arabidopsis* ATH1 Genome Array (GPL198) were chosen for this study as they provide extensive gene coverage and are widely used. GEO currently holds 1920 and 9106 samples and 114 and 709 series records (group of related samples) belonging to GPL2025 and GPL198 platforms, respectively. In total, we analyzed 305 and 220 samples of rice and *Arabidopsis*, respectively, belonging to 28 series records. The number of selected series, sample records and number of controls and treatments for each stress condition is given in Table S3.1. Complete list of selected series and sample records including their GEO IDs and brief description is given in Table S3.2.

3.3.2 Identification of differentially expressed genes

The raw intensity CEL files of the selected samples were downloaded from GEO and intensity values were extracted from the CEL files using the bioconductor package Affy in R (Gautier et al., 2004), quality checked using the package, ArrayQualityMetrics (Kauffmann et al., 2009) and the samples failing two or more of its quality tests were removed. The samples of each stress were normalized together using Robust Multichip Average (RMA) method (Irizarry et al., 2003). The probes were then matched to their loci based on annotation provided by array element mapping facility at TAIR portal for *Arabidopsis* (http://www.*Arabidopsis*.org/portals/expression/microarray/microarrayElementsV2.jsp) and at ricechip.org (http://www.ricechip.org) for rice. Probes with no match or ambiguously matching multiple loci were discarded. The retained probes and their normalized intensity values were then loaded into oneChannelGUI environment to perform non-specific filtering of probes with relatively small signal distribution using Inter Quartile Range (IQR) filter at most stringent setting (0.5) and probes with very low intensity values (probes below threshold log2(50)=5.64 in ≥90% of arrays). An example of resultant distribution of retained probes after filtering is shown in Supplemental Fig. S3.3.1.

Differentially expressed genes (DEGs) were identified using Rank Product method (Breitling et al., 2004). Rank Product is a non-parametric method returning up and down regulated genes, their fold change (FC), p-values and percentage of false predictions (PFP). It was shown to perform better than other methods including significance analysis of microarrays (SAM), Fisher's Inverse $\chi 2$ test and t-based hierarchical modeling (Hong and Breitling, 2008) and is widely used for meta-analysis studies combining data sets from different origins of the sample pool to increase the power of identification (Tseng et al., 2012). We used the function RPadvance of the bioconductor package RankProd (Hong et al., 2006) which is specifically designed for meta-analysis. The number of permutation tests was set to 250. The function topGene with a PFP cut-off value of ≤0.01 was used to output differentially expressed genes.

Among multiple probes matching the same locus, the probe ID with highest fold change was retained.

The orthologs between rice and *Arabidopsis* were obtained by parsing the gene families reported in GreenPhylDB (Rouard et al., 2011) which were identified based on analysis of complete proteomes of 16 plant species, cross referencing a number of resources (UniProtKB, Pubmed, InterPro, MEME motifs, KEGG pathways).

3.3.3 Functional enrichment analysis

Gene ontology analysis was carried out using the Singular Enrichment Analysis (SEA) tool offered by agriGO (Du et al., 2010) at default settings of Fisher t-test ($p<0.05$), False Discovery Rate (FDR) correction by Hochberg method and five minimum number of mapping entries against species specific pre-computed background reference. KEGG orthology (KO) terms associated with a gene correspond to KEGG pathway nodes and BRITE hierarchy nodes (Mao et al., 2005). To identify enzymes and proteins encoded by differently expressed genes and their associated metabolic and signaling pathways in each stress condition, we performed enrichment analysis of KO terms and determined the significance based on hypergeometric distribution p-values with <0.05 cut off value. Further analysis of biological pathways was carried out using the tool Database for Annotation, Visualization and Integrated Discovery (DAVID) v6.7 (Huang et al., 2009). Information on transcription factors (TFs) genes in rice and *Arabidopsis* was obtained from the database PlnTFDB (Perez-Rodriguez et al., 2010) and analyzed for enrichment of TF families in various gene sets.

3.3.4 Co-expression network analysis

To identify co-expression modules within SRGs, we extracted the normalized, log transformed gene expression values of each stress condition from the microarray experiments used in meta-analysis and performed Weighted Gene Co-expression Network Analysis (WGCNA) (Langfelder and Horvath, 2008). Briefly, WGCNA procedure calculates Pearson's correlation matrix for all genes, transforms the correlation matrix by raising all values to a power ß (soft thresholding as biological networks are

small world and scale free (Albert, 2007)), calculates a topological overlap matrix (TOM) from the transformed correlation matrix, converts the topological overlap matrix into a dissimilarity matrix, creates a hierarchical cluster tree based on the dissimilarity matrix, and identifies gene co-expression modules from the hierarchical cluster tree using a dynamic tree cut procedure. The blockwiseModules function of WGCNA package in R was used to generate the modules with powers 8, 6, 14 and 5 for RD, RB, AD and AB, respectively, which best approximate a scale-free topology (model fit >0.8) of the resultant network (Fig. S3.2). For this analysis, module size was 20-30, deep split was set at level 4 and tree merge cut height was 0.15-0.25. Heatmaps were constructed to depict the eigengenes from each identified module. Eigengenes represent a centroid measure of the expression levels of all genes in a cluster. The SRGs common to drought and bacterial stress were analyzed to find consensus modules showing co-expression patterns across stresses using the function blockwiseConsensusModules with the following settings: powers 7 and 10, minimum module size 30 and 15 for rice and *Arabidopsis*, respectively, with the merge cut height set at 0.15.

3.5 Results and discussion

3.5.1 Highly conserved expression status of genes common to drought and bacterial stresses

We identified a total of 5084 and 1618 DEGs referred herein as stress responsive genes (SRGs) in rice and *Arabidopsis*, respectively, combining the DEGs in drought and bacterial stresses together that were below FDR ≤0.01 (Fig. 3.1). Greater than 60% of genes were unique to individual stresses in all cases and AB (*Arabidopsis* Bacteria) had highest percent (~75%) of unique SRGs (799 genes). The number of up and downregulated SRGs are shown in Fig. S3.3A and complete list of genes along with their fold change values is given in Table S3.3. Among the 1214 SRGs common to the stresses studied in rice, majority of the genes were expressed in same direction (72% or 874) with 565 up and 309 downregulated in both drought and bacterial stresses. Similarly, higher

number of SRGs (170 out of 272 or 62.5%) common to both stresses studied in *Arabidopsis* was expressed in same direction with 93 and 77 genes up and downregulated, respectively. This finding elucidates that these set of genes and their associated biological processes are altered similarly as part of stress response in a wide range of tissues, developmental stages, stress levels and ecotypes (Table S3.2A). Among the genes with non-conserved expression pattern, the proportion of genes showing downregulation in drought and upregulation in bacterial stress (255 or 21% of 1214) was higher in rice while upregulation in drought and downregulation in bacterial stress (66 or 24% of 272) was higher in *Arabidopsis* (Fig. S3.3D).

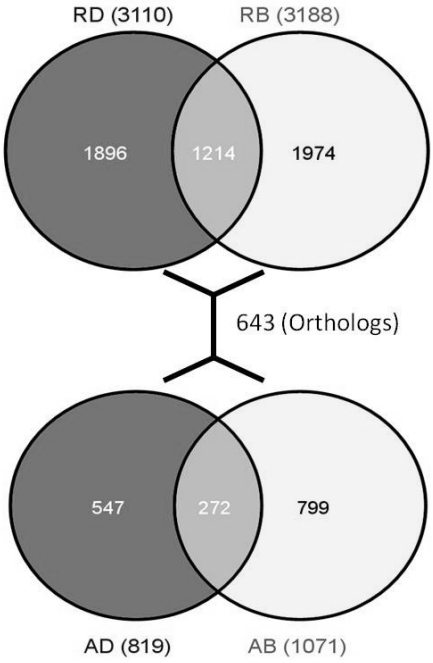

Figure 3.1: Number of unique and common differentially expressed genes (DEGs) found in rice and Arabidopsis. The number of orthologous genes found between rice and Arabidopsis DEGs are also shown. RD: Rice Drought, RB: Rice Bacteria, AD: Arabidopsis Drought, AB: Arabidopsis Bacteria.

The average fold change observed for SRGs was about 1.52, 0.93, 1.28 and 0.99 for RD (Rice Drought), RB (Rice Bacteria), AD (*Arabidopsis* Drought) and AB (*Arabidopsis* Bacteria) stresses respectively. The number of SRGs with fold change (FC) value ≥1.5 was higher in drought stress (51% and 26% in RD and AD respectively) and lower in bacterial stress (4% and 3% in RB and AB respectively), majority of which were

part of downregulated genes. Especially three genes showed >11 fold downregulation in RD, with LOC_Os05g47540 annotated as 'CPuORF26 - conserved peptide uORF-containing transcript, expressed' under expressed 20.86 folds. Upstream open reading frames (uORFs) are small open reading frames found in the 5' UTR of mature mRNA which regulate translation of major ORFs (mORFs) that code for transcription factors, signal transduction factors and developmental signal proteins (Hayden and Jorgensen, 2007). Multiple studies have reported the involvement of uORFs in translation repression of target genes in response to stress conditions (Jorgensen and Dorantes-Acosta, 2012). We found this gene to be downregulated also in RB (FC 1.55). LOC_Os10g36500 annotated as 'invertase/pectin methylesterase inhibitor family protein' is the second top DEG which was downregulated in both stress conditions (FC 11.34 and 1.20 in RD and RB, respectively). Pectin methylesterase inhibitors (PMEI) are invertase inhibitor-related defense proteins that play key roles in developmental transitions, wounding, senescence and abiotic stresses (An et al., 2008). Another gene that was highly downregulated in RD (FC 11.08) but upregulated in RB is LOC_Os04g39320 annotated as 'expressed protein'. In all the four stresses about 15-20% of SRGs were annotated as just 'expressed protein' or 'protein_coding' or 'unknown protein' (500, 556, 220, 157 DEGs in RD, RB, AD and AB, respectively) suggesting there are still hundreds of stress responsive genes with little or no functional information. We also found ~1% SRGs (27 and 34 genes in RD and RB, respectively) were annotated as retrotransposon related genes in rice. In *Arabidopsis*, 21 genes showed >4 fold downregulation under drought stress with AT1G22690 annotated as 'gibberellin-responsive protein' and AT5G03350, a legume lectin family protein showing 8.8 and 7.9 FC, respectively.

We found 643 orthologous genes between rice and *Arabidopsis* that are involved in stress response (Table S3.4). There were 255 orthologous genes differentially expressed in drought out of which 167 or 65% had their expression status conserved (73 and 94 were up and downregulated, respectively, in both rice and *Arabidopsis* genomes). Similarly, there were 280 orthologous genes differentially expressed in bacterial stress, out of which 211 or 75% had their expression status conserved. Majority of these were upregulated in both the genomes (134 or 63% SRGs). We also analyzed orthologs

between AD and RB, and found 72 SRGs with conserved upregulation. On the other hand, there were 102 SRGs with conserved downregulation between AB and RD (Fig. S3.4). There were 9 up and 8 downregulated orthologous genes found in all four stresses. One of these genes is a MYB TF that was highly downregulated, especially in drought (AT2G21650 (FC 3.5) and LOC_Os01G44390 (FC 5.5)). ARR6 and 7 (two-component response regulators) and their orthologous gene OsRR10 involved in cytokinin response system (Hwang and Sheen, 2001, Tsai et al., 2012) were also downregulated in all stresses. The upregulated genes in all four stress conditions include a NAC TF (AT1G69490 and LOC_Os03G21060), HAI-1 or highly ABA-induced PP2C gene 1 (AT5G59220 and LOC_Os05G38290) and a heavy metal-associated domain containing protein (AT5G52760 and LOC_Os10G38870). Expression of HAI-1 gene was shown to be induced by wound in *Arabidopsis* (Zhang et al., 2012a).

3.5.2 Functional enrichment analysis of SRGs

We found 623 unique GO terms enriched by SRGs in one or more stress conditions (Table S3.5). We analyzed gene sets that are up or downregulated separately for each stress as shown in Fig. S3.3B. Although the number of SRGs in *Arabidopsis* was only 1/3rd compared to those found in rice, total number of significant GO terms in *Arabidopsis* is close to rice reflecting the lack of annotation for a number of rice genes. Four way Venn diagram analysis revealed the number of GO terms common and exclusive to same stress (28 terms between RBU and ABU vs. 4 terms between RBU and ADU) and same species (68 between ADU and ABU vs. 10 between ADU and RDU) were higher than vice versa (Fig. S3.5). The top most significant GO term found in upregulated gene sets were response to water (FDR 5E-11), ribosome (2.9E-37), response to organic substance (3.4E-31) and response to biotic stimulus (2.4E-30) in RDU, RBU, ADU and ABU, respectively and in downregulated sets were catalytic activity (1.7E-24), photosynthesis (1.7E-16), thylakoid (2.6E-18) and response to chemical stimulus (1.3E-15) in RDD, RBD, ADD and ABD, respectively. The terms, ‘polysaccharide catabolic process’, ‘hydrolase activity, hydrolyzing O-glycosyl compounds’, ‘aromatic amino acid family metabolic process’, ‘regulation of gene expression’, ‘transcription factor activity’

were significantly enriched in upregulated gene sets, while 'photosynthesis', 'circadian rhythm', 'cofactor biosynthetic process', 'substrate-specific transmembrane transporter activity' were significantly enriched in downregulated gene sets (Fig. 3.2).

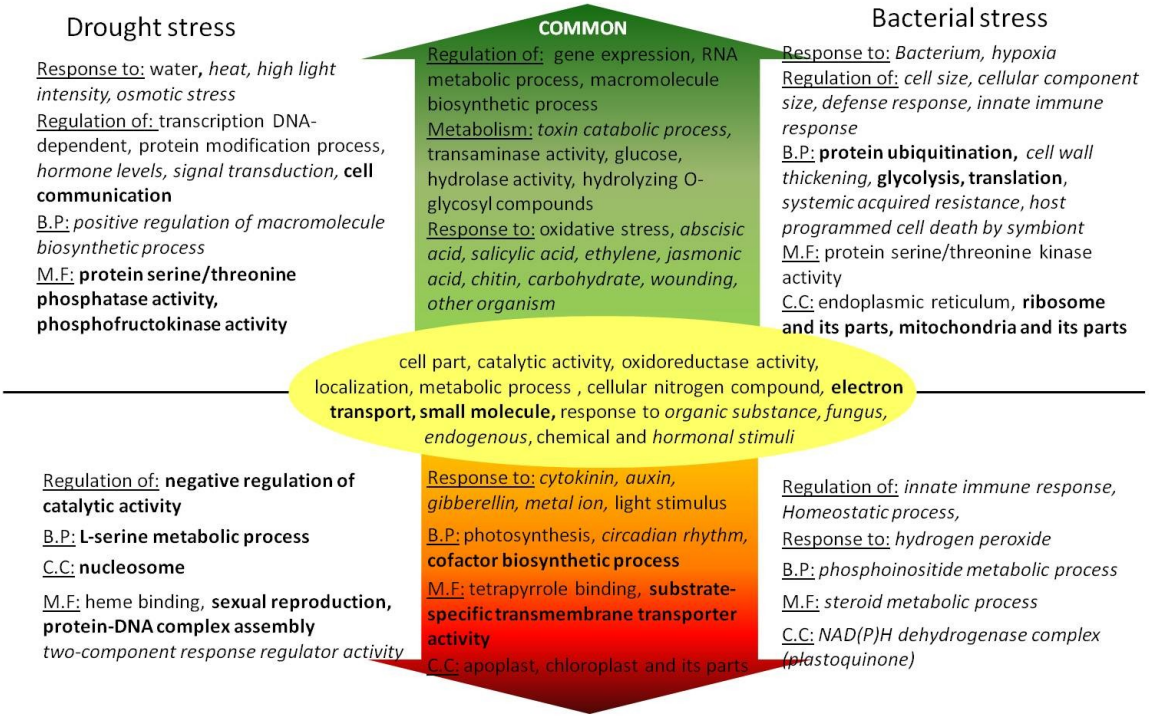

Figure 3.2: Summary of significant GO terms found in different stresses. Terms in green arrow indicate those that are commonly upregulated in drought and bacterial stress, and the terms besides the green arrow indicate those that are specifically upregulated in one stress. Similarly the terms in red arrow and those besides indicate the terms that are downregulated in both stresses and specific to one stress respectively. Terms in yellow oval were found both in up and downregulated gene sets. The terms in bold and those in italics are highly significantly found in rice and Arabidopsis respectively. B.P: Biological Process, M.F: Molecular Function, C.C: Cellular Component.

Terms related to hormones and their related functions showed clear distinction between the processes that are up or downregulated in a stress response especially in *Arabidopsis*. While terms related to the hormones auxins, cytokinins and gibberellins were downregulated, abscisic acid, salicylic acid, ethylene and jasmonic acid was upregulated both in drought and bacterial stresses. Abscisic acid (ABA) is known to play a central role in abiotic stress response by inducing stomatal closure resulting in reduction of transpiration (Pantin et al., 2013), regulating root growth, ion channels and gene expression (Duan et al., 2013). Further, it was found that ABA can have both

positive and negative effect on biotic stress signaling (Melotto et al., 2006). For example, ABA-induced stomatal closure prevented invasion of microbes through open stomata. Thus, recent findings increasingly suggest ABA as a key player in fine-tuning of cross talk between abiotic and biotic stress responses and therefore ABA production can be the crucial factor determining how well a plant responds to multiple stresses (Atkinson and Urwin, 2012). While the terms 'response to ethylene stimulus' and 'response to salicylic acid stimulus' were found both in ADU and ABU, the terms 'ethylene mediated signaling pathway' and 'salicylic acid mediated signaling pathway' were significant only in ABU, which is in agreement with their known functional roles in defense against pathogens and senescence (Vlot et al., 2009, Wilkinson et al., 2012). Further, the terms 'host programmed cell death induced by symbiont' and 'systemic acquired resistance (SAR)' mechanisms that are induced by salicylic acid were also found only in ABU. On the other hand, jasmonic acid biosynthetic process was significant only in ADU although jasmonic acid mediated signaling pathway was significant both in ADU and ABU. Jasmonic acid (JA) plays a key role in defense response especially against necrotrophic pathogens and wounding acting antagonistically to salicylic acid which is majorly involved in resistance to biotrophic pathogens (Thaler et al., 2012). JA also has a role in the formation of antioxidants that regulate ascorbate and glutathione metabolism (Brossa et al., 2011) explaining our observation of its increased synthesis in drought stress. The downregulation of all of the major plant growth and development promoting hormones such as auxins, cytokinins and gibberellins across diverse stress conditions indicates various processes including cell differentiation, chloroplast biogenesis, flowering and reproduction (Bari and Jones, 2009, Cui and Luan, 2012), controlled by them are pushed to backseat while processes related to reprogramming of metabolism, gene expression, balancing of homeostasis and modulation of defense and immunity are given higher priority. The above observations are further supported by a number of terms related to photosynthesis and biosynthesis of its components including 'chloroplast', 'photosystem', 'photosynthetic membrane', 'photosynthesis, light reaction', 'photosynthetic electron transport chain' that were highly enriched in all four downregulated gene sets but none of the upregulated gene sets.

GO terms related to various metabolic processes including carbohydrates, amino acids, proteins, ribosomes, translation and nucleobases were significantly enriched in RBU. Translation is a highly energy expensive process and its regulation via protein phosphorylation, initiation factor isoforms, RNA sequence element interactions, and small RNAs enable cells to rapidly and reversibly control gene expression in response to environmental changes (Muench et al., 2012). Upregulation of a number of translation related GO terms in rice under bacterial stress suggests cellular adjustments at translational level upon bacterial infection. The term 'response to water' was highly enriched in RDU (FDR 5E-11) and ADU (FDR 4.6E-19) and the term 'response to water deprivation' was highly enriched in ADU (1.5E-18). A number of terms related to regulation of gene expression and metabolic processes including 'transcription factor activity', 'nucleic acid metabolism', and 'chitin catabolic process' were enriched in three or all of the upregulated gene sets. Both positive and negative regulation of response to stimulus was found in ABU. The term 'negative regulation of defense response' was also significantly enriched in ABU (FDR 8.5E-05). The SRGs associated with the above GO term, EDS1 (Enhanced disease susceptibility 1) and PAD4 (Phytoalexin deficient 4) directly interact and induce salicylic acid biosynthesis in response to biotrophic pathogens (Shah, 2003). A mutant of EDS1 was found to be disease resistant (Frye et al., 2001).

The enriched KEGG orthology (KO) terms in different SRG sets revealed many similar patterns as that of GO analysis that can be seen by the top KO terms and their associated pathways in Table S3.6. Enrichment of 'jasmonate ZIM domain-containing' proteins (JAZs) and 'auxin responsive GH3 gene family' proteins in the upregulated SRGs of *Arabidopsis* substantiate recent findings that these proteins negatively regulate downstream processes of hormonal activity especially those related to plant growth and development (Park et al., 2007, Chung et al., 2008, Cheng et al., 2011). On the other hand, KO terms, 'two-component response regulator ARR-A family' involved in negative regulation of cytokinin signaling via phospho relay (To et al., 2007) and 'SAUR family proteins' which are primary auxin-inducible genes involved in auxin transport and organ elongation (Chae et al., 2012) were enriched in downregulated gene sets of both the

stresses. Reactive oxygen species (ROS) have been proposed as a central component of plant adaptation to both biotic and abiotic stresses (Dat et al., 2000). Glutathione S-transferase (GST) plays a key role in scavenging ROS and detoxification and is differentially activated by stress-induced plant growth regulators (Moons, 2005). GST was upregulated in both the stresses and was also part of ADD.

A number of terms related to enzymes involved in biosynthetic pathways of amino acids including 'peroxidase', 'tyrosine aminotransferase' and 'serine O-acetyltransferase' were part of downregulated gene sets (Table S3.7). The KO term 'Cellulose synthase A (CesA)' was highly enriched in RDD. Several studies reported disruption of genes involved in biosynthesis of cellulose enhanced stress tolerance (Chen et al., 2005, Hernandez-Blanco et al., 2007, Song et al., 2013). As also revealed by GO analysis, the term 'small subunit ribosomal protein S4e' was enriched in RBU and 'ferredoxin' involved in photosynthesis was enriched in RBD. Heat shock protein 70 (Hsp70) is one of the most abundant heat shock proteins in eukaryotic cells which bind to hydrophobic patches of partially unfolded proteins preventing protein aggregation (Mayer and Bukau, 2005). Hsp70 was enriched in both the upregulated gene sets of rice.

The KEGG pathways found significant by the tool DAVID with p-value <0.05 in SRG sets are shown in Fig. S3.6. The pathway 'fatty acid metabolism' was enriched both in RDU and RBU. Plants acclimating to stress modulate membrane fluidity and levels of oleic acid and linolenic acid using lipases facilitating proper functioning of critical integral proteins during stress (Upchurch, 2008). α-linolenic acid released under stress from chloroplast membranes is a major parent compound for an array of messenger compounds derived via oxidative modification by ROS (Demmig-Adams et al., 2013) including jasmonic acid (Staswick, 2008, Gfeller et al., 2010). The pathway 'α-linolenic acid metabolism' was highly significant in ADU and RBU. A number of pathways related to biosynthesis of secondary metabolites were enriched in upregulated sets including biosynthesis of alkaloids from shikimates, purines, histidine, terpenoids and polyketides. Phenylpropanoids, derived from a very limited set of core structures of shikimate pathway are modified by oxygenases, ligases, oxidoreductases and transferases to generate an enormous number of secondary metabolites (>200,000) including lignins,

suberins and tannins which contribute substantially to the robustness of plants facing stress (Vogt, 2010) and are also implicated in providing nutritional and medicinal benefits for animals and humans due to their potent antioxidant activity (Tohge et al., 2013). Phenylpropanoid biosynthesis was enriched in drought especially in rice but was found both in up and downregulated gene sets suggesting differential regulation of the enzymes resulting in synthesis of different end-products. The biosynthetic pathway of flavonoids from phenylpropanoid derivatives was enriched in ABD.

Biosynthesis and metabolic pathways of aromatic amino acids, phenylalanine, tyrosine and tryptophan and degradation pathways of lysine, valine, leucine and isoleucine were enriched in upregulated gene sets (ABU, ADU, RBU and RDU). The aromatic amino acids are also synthesized via the shikimate pathway playing crucial roles in plant growth, development, reproduction, defense, and environmental responses (Tzin and Galili, 2010, Maeda and Dudareva, 2012). Recent reports indicate reduction in starch biosynthesis and accumulation, and increased consumption of storage substances under drought (Harb et al., 2010, Lenka et al., 2011) resulting in elevated levels of hexose sugars (glucose and fructose) (Shu et al., 2011a). Our analysis revealed upregulation of starch and sucrose metabolism, glycolysis/gluconeogenesis and pentose phosphate pathway in both drought and bacterial stresses. As observed in GO analysis, a number of pathways related to photosynthesis were enriched in downregulated gene sets including porphyrin and chlorophyll metabolism, carbon fixation in photosynthetic organisms and carotenoid biosynthesis. Similar to the observation of enrichment of GO term 'nucleobase, nucleoside, nucleotide and nucleic acid biosynthetic process' in RDD, 'amino sugar and nucleotide sugar metabolism' pathway was also enriched in RDD.

Out of the 82 and 56 known TF families/regulators, 34 (41%) and 38 (67%) were found in one or more gene sets of *Arabidopsis* and rice, respectively (Fig. S3.3C). A comparative list of the number of TFs belonging to each TF family found in different stresses is given in Table S3.8. Among the large TF families, higher numbers of NAC, ERF, AP2-EREBP and C2H2 family members were found in upregulated gene sets while higher numbers of bHLH and MYB_related family members were found in downregulated gene sets. WRKY TFs were the highest in the upregulated set of bacterial

stress both in rice and *Arabidopsis*. WRKY TFs are considered to be at the heart of global regulation of plant immunity by modulating its immediate downstream target genes which include MAP kinases and other TFs (Pandey and Somssich, 2009). 'CO-like' TF family members were the highest in RBU (17 TFs) and RDD (11) but low in RBD (1) and RDU (3) indicating an inverse expression relationship between drought and bacterial stress. *CO* (*CONSTANS*) gene and other members of CO-like TF family play an important role in regulation of flowering and act between the circadian clock and genes controlling meristem identity (Griffiths et al., 2003). A high number of HSF (heat shock transcription factor) family members were found in upregulated gene sets of rice. Seven HD-ZIP (heomodomain leucine zipper motif) members were found in RDU only. Out of 16 Tify family members in *Arabidopsis*, seven were found in ADU. Tify is a novel TF family with JAZ motifs, is implicated to play a critical role in jasmonate signaling pathway (Chung and Howe, 2009). Members of this family were reported to be strongly induced under drought conferring improved tolerance to drought and high salinity (Ye et al., 2009).

3.5.3 Gene network analysis revealed tightly co-expressed modules of SRG sets

Gene coexpression networks, built using a set of microarray samples as input, can help elucidate tightly coexpressed modules that are a mixture of genes with known and unknown functions, identify hub genes, and candidate genes which can be used as biomarkers (Ficklin et al., 2010, Allen et al., 2012). Using Weighted Gene Co-expression Network Analysis (WGCNA), we divided SRGs into 11, 10, 5, 8 modules of RD, RB, AD and AB, respectively, excluding a grey color module listing genes that did not significantly co-express with any other group of genes. The module of each SRG indicated by module color, k_{IM} (intramodular connectivity), a measure of how well connected or co-expressed a given gene is, with respect to other genes in its module, MM (Module Membership), a measure of module membership correlating its gene expression profile with the Module Eigengene (ME, which is the first principal component of a given module also considered as a representative of the gene expression

profile of the module) (Langfelder and Horvath, 2008) and p-values are given in Table S3.3.

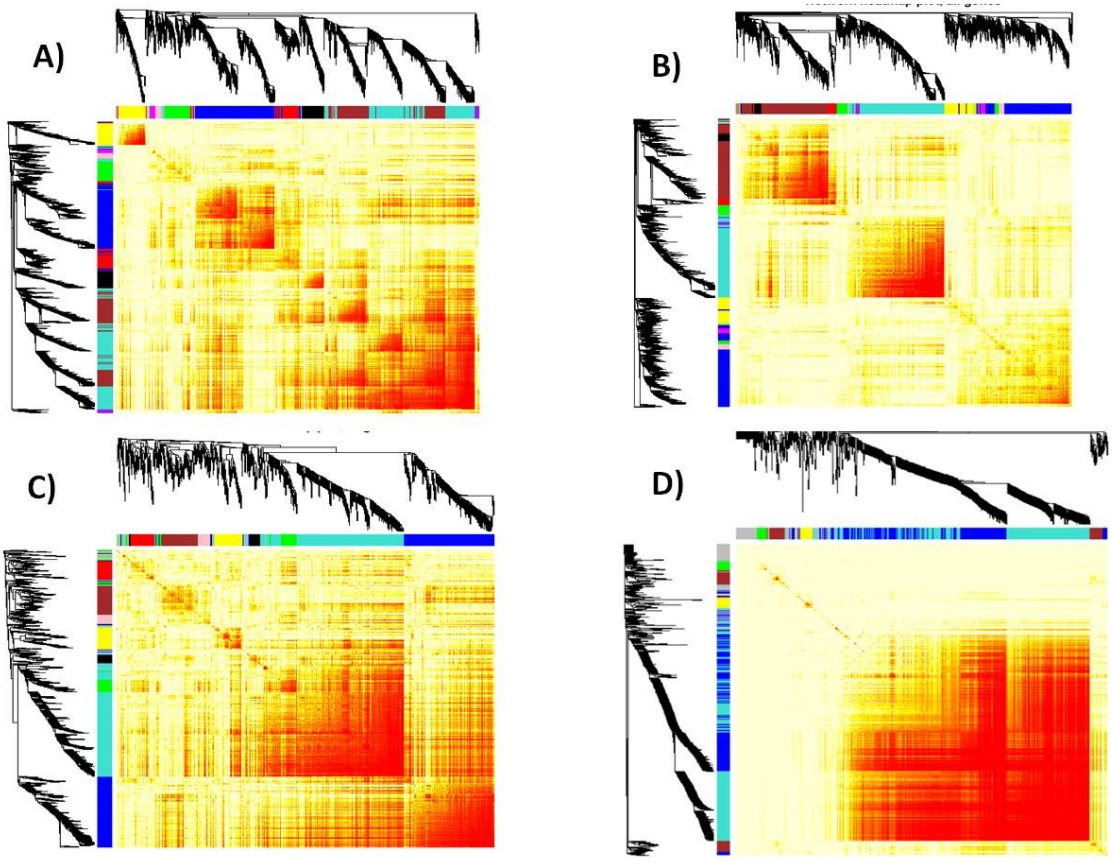

Figure 3.3: Dendrograms and heatmaps of SRGs divided into tightly co-expressed modules by the R statistical package WGCNA. A) RB, B) RD, C) AB, and D) AD. The DEGs were clustered based on co-expression patterns as represented by the dendrogram and correlation heat map. Clusters of like-regulated genes are referred to as modules and are indicated by different colors. Grey color represents the genes that could not be assigned to a module. Intensity of red coloring in the heat map indicates strength of correlation between pairs of genes on a linear scale.

The long length of the dendogram branches and corresponding intense red color in the heat maps of co-expression modules illustrate high co-expression of SRGs within modules and less co-expression outside the module (Fig. 3.3). We used unsigned correlations so that positively and negatively correlated genes could be grouped into the same module. Yet, a number of modules showed high enrichment of either up or down regulated genes (Table S3.9). For example, the largest module (turquoise) found in RD

with 846 SRGs was made up of 663 (78%) downregulated genes while the second largest module (blue) with 763 SRGs was made up of 618 (79%) upregulated genes. We compared the 11 RD modules detected by us against 15 drought-responsive modules of rice found by another recent study using Markov Cluster (MCL) algorithm (Zhang et al., 2012b). Out of those 15 modules, 14 modules were made up of 28-75% of RD SRGs, most of which significantly overlapped with one of the RD modules. For example, module 2 found by Zhang et al (Zhang et al., 2012b) was made up of 213 genes, out of which 146 (68.5%) were part of SRGs and 116 (90%) of those overlapped with RD turquoise module. The module eigengene (ME) of the RD turquoise module has low values in all drought arrays compared to control indicating that most of the genes are downregulated (green color in the heatmap) (Fig. 3.4A). The top functional terms enriched in this module were predominantly related to photosynthesis. In the blue module, ME has higher values in all drought arrays compared to control indicating that most of the genes are upregulated (red color in the heatmap) (Fig. 3.4B). The top functional term of blue module was 'response to water' followed by protein domains 'dehydrin' and 'LEA'. Late embryogenesis abundant (LEA) proteins are extremely hydrophilic proteins implicated in desiccation tolerance and stabilization of proteins and membranes during drying (Hand et al., 2011). The blue module had a very high number of TFs than turquoise (64 compared to 38 TFs) (Table S3.10) although it was made up of less number of genes than turquoise. Majority of blue module TFs were from ERF and NAC families while turquoise had higher number of bZIP and CO-like TFs.

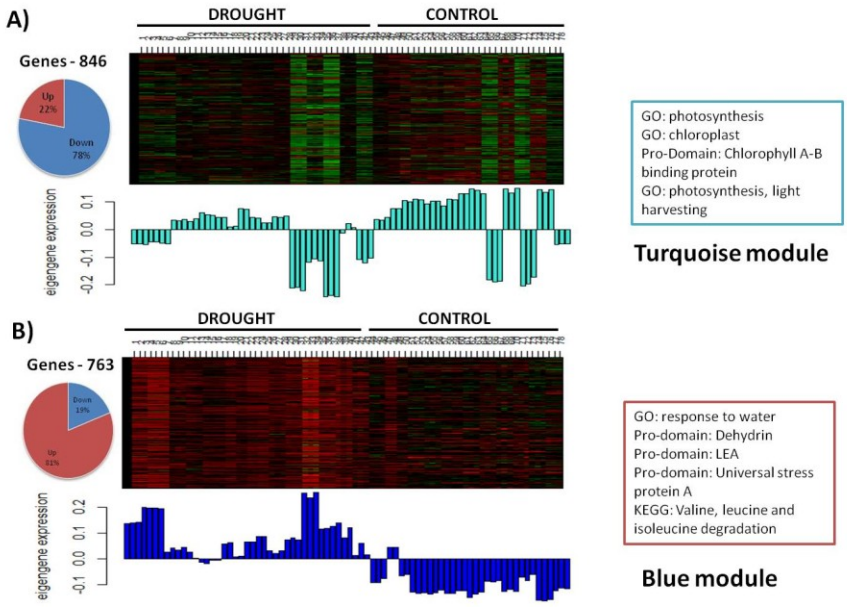

Figure 3.4: Heatmaps of turquoise and blue modules in rice under drought stress. The x-axis represents microarray samples grouped into drought treated and control samples and y-axis represents genes found in the module. Below the heatmap the corresponding module eigengene expression values are shown. The most significant functional terms found in the module are also shown. The number of genes found in each module and the percentage of up and downregulated genes in each module are shown as a pie chart.

Functional enrichment analysis of each of the co-expression modules revealed a number of significant terms with FDR <0.05 (Table S3.11) especially in *Arabidopsis* which were proportional to their module size. However, in rice, there was large variation in number of significant functional terms compared to module size (Table S3.9). For instance, the RD module brown (size 732) had 83 significant terms but blue (size 763) had only 8 terms with FDR <0.05. Further analysis of these modules revealed higher number of genes annotated as 'expressed protein', 'DUF – Domain of unknown function' in blue module (129, 26 and 260) compared to brown module (96, 9 and 211). There were 51 and 278 genes in blue and brown modules, respectively, with high intramodular connectivity (k_{IM} value >100), out of which 11 and 31 genes were annotated as 'expressed protein' in blue and brown modules, respectively. These genes would be important candidates for further investigation as they might be playing significant role in stress response.

Under same stress, a number of modules in rice and *Arabidopsis* showed relatedness in functionality, indicating conservation of co-expression of functionally related genes across species. The module AD turquoise was related to RD brown with shared terms, response to oxidative stress (GO:0006979, AD FDR=5.18E-07, RD FDR=5.32E-07) and calcium ion binding. The module AD blue was related to RD turquoise with terms photosynthesis (GO:0015979, AD FDR=9.24E-07, RD FDR=2.0E-19) and other similar terms. AB yellow was related to RB magenta with shared terms, 'aromatic compound biosynthetic process' and 'cellular amino acid biosynthetic process'. RB red with 203 upregulated genes out of 206, had 26 TFs which is double the percent of TFs found in other modules. Most of the TFs in this module belong to WRKY and MYB families with the top gene being a MYB TF, LOC_Os04g43680. The only downregulated genes in this module are LOC_Os05g37820 (major facilitator family transporter), LOC_Os09g35010 (dehydration-responsive element-binding protein) and LOC_Os02g51910 (cytokinin-O-glucosyltransferase 2).

Among the modules found in AD, brown (size 64) had 63 upregulated genes and 22 TF genes (34%), and showed enrichment of 44 functional terms including response to various hormones and endogenous stimuli like water deprivation, salt, cold, temperature and chitin. There were 6 TFs including WRKY33 and WRKY40 in the top 10 genes in this module based on k_{IM} values. Among AB modules, yellow module with 90% (72 out of 80) downregulated genes and 18.75% of TFs showed enrichment of a number of terms related to secondary metabolic process including biosynthesis of aromatic compounds, flavonoids and phenylpropanoids.

3.5.4 Consensus co-expression modules of drought and bacterial stresses

The expression profiles of the SRGs common to drought and bacterial stresses was utilized to detect consensus modules that would reveal sets of genes with similar co-expression patterns in both the stresses. We found 9 and 4 consensus modules (excluding grey module for genes that did not co-express with others) based on 1214 and 272 SRGs differentially expressed both in drought and bacterial stress in rice and *Arabidopsis*, respectively (Fig. 3.5 and table 3.1). The color coded tables below the dendrograms in

Fig. 3.5 show the correspondence between consensus modules and modules found individually in drought and bacterial stress revealing several of the modules with preserved module structure. Consensus modules brown, turquoise and blue in rice and turquoise and brown in *Arabidopsis* showed significant overlap with their counterparts indicating the module structure in drought and bacterial stress to be very similar. A complete list of SRGs with their consensus modules and k_{ME} values which is a measure of module membership by correlating its gene expression profile with its module eigengene is given in Table S3.12.

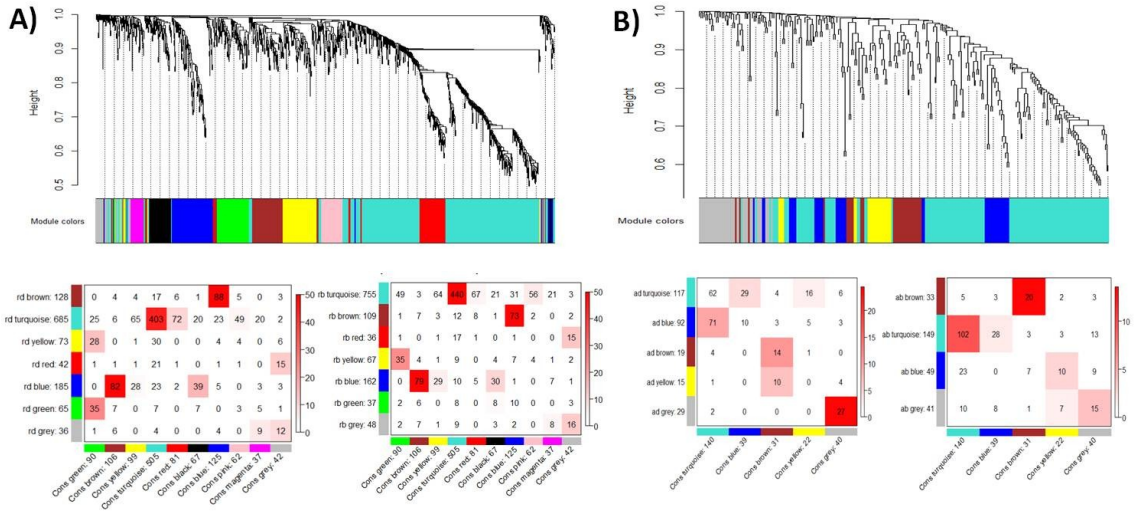

Figure 3.5: Clustering dendrogram of genes and consensus modules found in A) rice and B) Arabidopsis. The correspondence between consensus modules and modules found individually in drought and bacterial stress based on the expression values of the common genes are also shown as a table. Each row of the table corresponds to individual stress specific module (labeled by color as well as text along with the number of genes in the module), and each column corresponds to one consensus module. Numbers in the table indicate gene counts in the intersection of the corresponding modules. Coloring of the table encodes -log(p), with p being the Fisher's exact test p-value for the overlap of the two modules. The stronger red color indicates more significant overlap.

Among the 9 consensus modules found in rice, three modules showed conservation of differential expression in >90% of genes. Of these, module red contains majorly downregulated genes while brown contains upregulated genes. Red module was enriched with terms ‘ribonucleoprotein’ and ‘rotamase’ and brown was enriched with terms ‘valine, leucine and isoleucine degradation’ and ‘NAM protein’. Interestingly, two modules (magenta and black) showed >92% of genes with reversed expression status

suggesting that these set of genes possibly play a co-ordinated role specific to the stress condition. Most of the genes in these modules were downregulated under drought but upregulated under bacterial stress elucidating the differences in abiotic and biotic stress responses. We further investigated if this trend can be observed in other stresses, using the tool Genevestigator (Zimmermann et al., 2004). Analysis of the expression profile of genes in magenta module under salt (3 microarray studies) and fungal (3 studies and 2 pathogens, *B. graminis* and *M. oryzae*) stress conditions identified most of the genes to be highly up and downregulated under fungal and salt stresses, respectively (Fig. 3.6). Magenta color module showed enrichment of GO terms 'electron transport' and 'oxidoreductase activity' and black was significantly enriched in the following protein domains: 'Glycoside hydrolase, chitinase active site', 'DNA-binding WRKY', 'Bet v I allergen' and 'VQ'. Bet v 1 belongs to plant pathogenesis-related proteins (PR-10) family that is involved in plant development and defense systems via interactions with plant hormones (Fernandes et al., 2013). VQ is a small motif found only in plants. A recent study has shown that VQ motif containing proteins act as co-activators of WRKY33 in *Arabidopsis* as part of plant defense response (Lai et al., 2011, Cheng et al., 2012). The gene LOC_Os01g61080 (WRKY24) which is an ortholog of WRKY33 of *Arabidopsis* was also part of black module. Occurrence of VQ motif containing genes (LOC_Os05g44270 and LOC_Os03g20440) and WRKY24 in the same module and upregulation of all three under bacterial stress and downregulation under drought stress suggests these genes play a similar role in rice.

Table 3.1. List of consensus co-expression modules found in each stress gene set

Rice Modules	Module Size	No. of TFs	RD Down/Up	RB Down/Up	Conservation of gene expression (%)
Brown	106	10	8/98	4/102	94.34
Red	81	2	72/9	75/6	93.83
Yellow	99	8	32/67	39/60	90.91
Turquoise	505	19	396/109	363/142	82.38
Blue	125	5	114/11	88/37	72.80
Green	90	5	17/73	33/57	48.89
Pink	62	2	57/5	28/34	46.77
Black	67	7	62/5	2/65	7.46
Magenta	37	3	37/0	1/36	2.70
Arabidopsis Modules			**AD Down/Up**	**AB Down/Up**	
Turquoise	140	14	63/77	98/42	53.57
Blue	39	6	22/17	20/19	64.10
Brown	31	7	2/29	1/30	90.32
Yellow	22	0	7/15	0/22	68.18

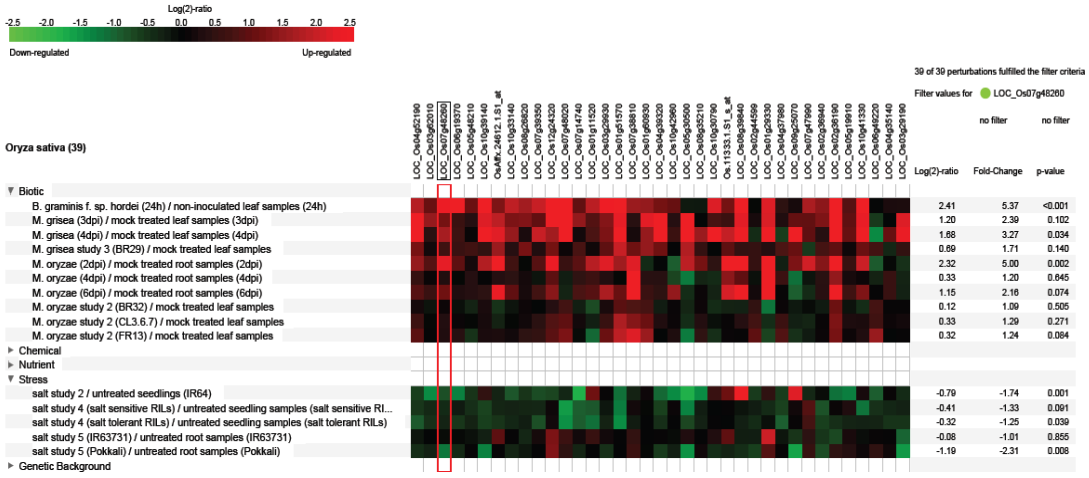

Figure 3.6: Gene expression profile of rice consensus module magenta under fungal and salt stresses using the tool Genevestigator. The heatmap shows color coded values based on Log(2)-ratio of test/control samples in different studies. A brief description of test/control samples including tissue, treatment and cultivar is also given. The top TF gene WRKY47 (LOC_Os07g48260) is highlighted with a red box and the corresponding log(2)-ratio, fold change and p-values across different microarray studies are shown.

Among the *Arabidopsis* consensus modules, brown and yellow were made up of mostly upregulated genes. In brown module, 28 (90%) out of 31 SRGs were upregulated in both the stresses. It contained four WRKY TF genes including WRKY33 (AT2G38470) which was also found in rice consensus module black. The top three SRGs of brown module based on k_{IM} values are AT3G23250 (MYB15), AT2G22880 (VQ motif-containing protein) and AT3G25780 (allene oxide cyclase 3), which is one of the enzymes involved in jasmonic acid bioysnthesis. The top three SRGs found in yellow module are AT5G67340 (armadillo/beta-catenin repeat family protein) which functions in ubiquitin-protein ligase activity, AT4G01700 (chitinase family protein) and AT5G50200 (wound-responsive gene 3), which encodes a high-affinity nitrate transporter.

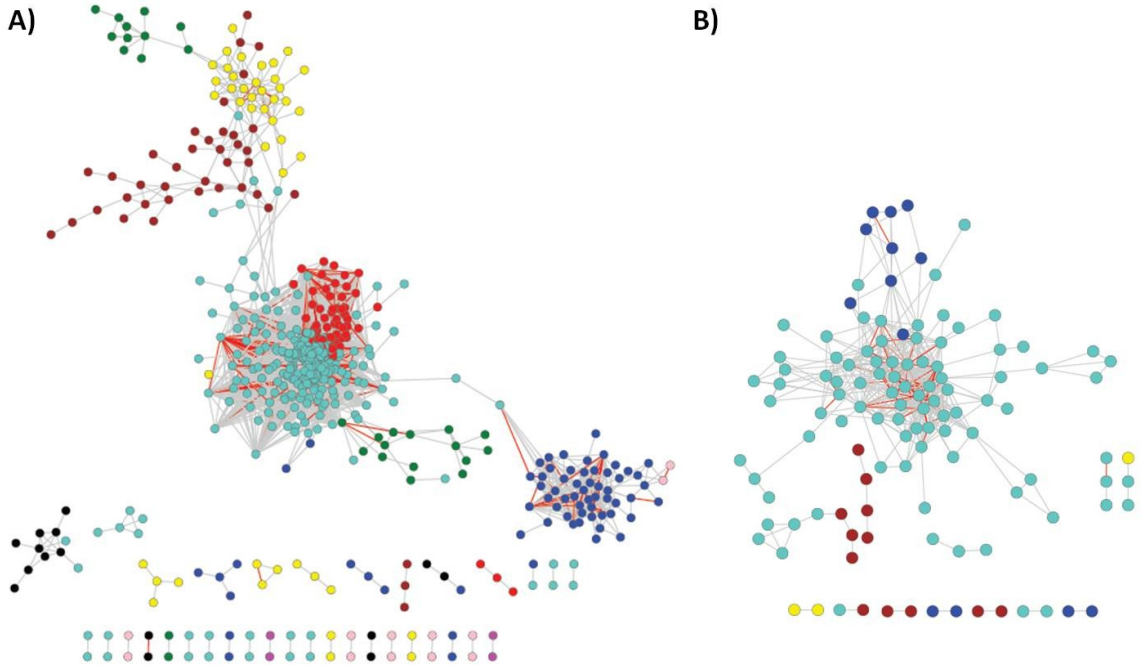

Figure 3.7: Coexpression network of SRGs common to drought and bacterial stresses. (A) rice (B) Arabidopsis. Nodes are color coded based on consensus modules found by WGCNA. Edges are constructed between genes with correlation coefficient (r) >0.8. The edges with r >0.8 are show in red.

We further analyzed the consensus co-expression modules by constructing a network based on co-expressed genes with high absolute Pearson correlation coefficient

(r >0.8) in both drought and bacterial stresses. There were 16,576 edges between 585 co-expressed genes in rice (Fig. 3.7A and Table S3.13A). One of the top edges was between LOC_Os02g43790, an ethylene-responsive TF and LOC_Os02g41510, a MYB TF with r >0.98 in both stresses. Color coding of nodes in network with their consensus module color showed clear grouping of genes from the same module with high number of intra-modular edges. For example, majority of blue consensus module genes had edges within the group and was largely isolated from all other modules. This indicates that these set of genes are co-regulated and exhibit stress specific co-functionality. Gene ontology analysis revealed enrichment of terms 'cytoplasmic membrane-bounded vesicle' (FDR=0.003) and 'endopeptidase activity' (FDR=0.02). Interestingly, these blue module genes were connected to the largest module (turquoise) via only one gene, LOC_Os05g09724 a HAD (haloacid dehalogenase) superfamily phosphatase which are involved in diverse housekeeping and secondary metabolism activities (Allen and Dunaway-Mariano, 2009). Red module showed the highest percent of genes (71 out of 81 or 87.6%) with a number of edges having r >0.8 in both the stresses. Black module had 15 genes (out of 67 or 22%) including 5 TFs with edges showing r >0.8, all of which showing non-conserved expression status between drought and bacterial stresses. In *Arabidopsis*, there were 509 edges between 119 genes showing r >0.8 in both stresses. Color coding the nodes with consensus module colors revealed that most of the edges were between genes of turquoise module (Fig. 3.7B and Table S3.13B). The top most co-expressed genes were AT3G51420 (Strictosidine synthase-like 4) and AT1G70760 (Chlororespiratory reduction 23) with r >0.98 in both the stresses.

3.6 Conclusion

In this study, we performed meta-analysis of microarray studies and identified differentially expressed genes in rice and *Arabidopsis* from a wide variety of samples under drought and bacterial stresses. This type of approach enhances sensitivity in the identification of important stress response genes that could be missed by studies that are limited to specific tissue or developmental stage or level of stress. Comparative analysis of the DEGs identified common stress responsive genes between stresses and across

species. Functional enrichment analysis revealed the biological processes, cellular pathways and transcription factor families that are commonly and exclusively altered under different stresses. The knowledge gained in this study on various molecular mechanisms like biosynthesis of secondary metabolites and stress specific roles of plant hormones vastly adds on to our understanding of stress response and its regulation. Weighted gene co-expression network analysis divided genes into individual and consensus modules and revealed sets of genes with conserved and reversed expression status. A number of genes with high connectivity, conserved expression but with poor annotation were identified. We propose these genes as potential candidates for stress response engineering.

Chapter 4: Identification and evaluation of stress responsive genes to distinguish multiple stress conditions in rice using machine learning approaches

Rafi Shaik and Wusirika Ramakrishna

The material contained in this chapter is under preparation to be submitted to the journal Plant Physiology

4.1 Abstract

Plant stress responses, broadly categorized into abiotic and biotic stresses are traditionally thought to be regulated by discrete signaling mechanisms. However, recent experimental evidence revealed a more complex picture where these mechanisms are highly entangled and can have synergistic and antagonistic effects on each other. In the present study, to comprehensively identify the shared stress responsive genes between abiotic and biotic stresses in rice, we performed meta-analyses of microarray studies from multiple abiotic and biotic stresses separately and found a list of 1377 common Differentially Expressed Genes (DEGs). About 70% of these common DEGs showed conserved expression status and majority of the rest (~21%) were downregulated in abiotic stresses and upregulated in biotic stresses. Using dimension reduction techniques, Principal Component Analysis (PCA) and Partial least squares Discriminant Analysis (PLS-DA), we were able to segregate abiotic and biotic stresses into two separate entities. The supervised machine learning model, Recursive-Support Vector Machine (R-SVM) showed that abiotic and biotic stresses can be classified with 100% accuracy using only 540 of the shared stress responsive genes. Further, using Random Forests (RF) decision tree model, we were able to classify 8 out of 10 different stress conditions with high accuracy. Comparison of lists of genes contributing most to the accurate classification by PLS-DA, R-SVM and RF revealed 196 common genes with a dynamic range of expression levels in multiple stress conditions. Functional enrichment and co-expression network analysis revealed the different roles of phytohormones and transcription factors in conserved and non-conserved gene-sets in regulation of stress responses. We envisage the top ranked genes identified in this study which highly discriminate abiotic and biotic stresses as key components to further our understanding of the inherently complex nature of multiple stress response in plants.

4.2 Introduction

With declining environmental conditions and scarce natural resources, the need to breed robust and high productivity crops is more important than ever. According to estimates, the world food productivity should be raised by as much as 70-100% to meet the energy needs of the world population which is expected to rise to 9 billion by 2050 (Godfray et al., 2010, Lutz and Samir, 2010). Rice is both a major food crop accounting for 20% of daily calorie intake of about 3.5 billion people (IRRI), and a model organism which shares extensive synteny and collinearity with other grasses. Thus, development of rice that can sustain a wide variety of adverse conditions is vital to meet the imminent global energy demands.

A broad range of stress factors divided into two major categories namely abiotic stresses encompassing variety of unfavorable environmental conditions such as drought, submergence, salinity, heavy metal contamination or nutrient deficiency and, biotic stresses caused by infectious living organisms such as bacteria, virus, fungi or nematodes negatively affect productivity and survival of plants. Advancements in whole genome transcriptome analysis techniques like microarrays and RNA-seq have revolutionized the identification of changes in gene expression in plants under stress, making it possible now to chart out individual stress specific biomolecular networks and signaling pathways. However, in field conditions, plants are often subjected to multiple stresses simultaneously, requiring efficient molecular mechanisms to perceive multitude of signals and to elicit a tailored response (Sharma et al., 2013). Increasing evidence from experimental studies suggests that the cross talk between individual stress-response signaling pathways via key regulatory molecules, resulting in the dynamic modulation of downstream effectors' is at the heart of multiple stress tolerance. A number of studies have identified many genes especially transcription factors and hormone response factors that play central role in multiple stresses and manifest a signature expression, specific to the stress condition. For example, ABA response factors are upregulated in majority of abiotic stresses activating an oxidative response to protect cells from ROS damage but were found to be downregulated in a number of biotic conditions possibly suppressed by immune response molecules (Cao et al., 2011).

The wide range of abiotic and biotic stress factors and numerous combinations of them in natural conditions, generating a customized stress response suggests identification and characterization of key genes and their co-expression partners which show an expression profile that discriminates abiotic and biotic stress responses would increase our understanding of plant stress response manifold and provide targets for manipulations that improve the stress tolerance of important food and energy crops. The availability of multiple genome-wide transcriptome data sets for same stress condition provides an opportunity to identify, compare and contrast stress specific gene expression profile of one stress condition with other stresses. Meta-analysis by combining similar studies provides a robust statistical framework to reevaluate the original findings, improve sensitivity with increased sample size and to test new hypotheses. Meta-analysis of microarray studies is widely used especially in clinical research to improve statistical robustness and detect weak signals (Liu et al., 2013, Rung and Brazma, 2013). For instance, thousands of samples belonging to hundreds of cancer types were combined which provided new insights into the general and specific transcriptional patterns of tumors (Lukk et al., 2010). Microarray studies are burdened with high dimensionality of feature space also called as 'curse of dimensionality' i.e. availability of very many variables (genes) for very few observations (samples). Machine learning algorithms (supervised and unsupervised) such as Principal Component Analysis (PCA), decision trees and Support Vector Machines (SVM) provide a way to efficiently classify two or more classes of data. Further feature selection procedures like Recursive-SVM (R-SVM) provide means to identify the top features contributing most to the accuracy of classification.

In the present study, we performed meta-analysis of stress response studies in rice using publically available microarray gene expression data conducted on a single platform (Affymetrix RiceArray). Meta-analysis of abiotic and biotic stresses was performed separately to identify differentially expressed genes involved in multiple stress conditions. The lists of abiotic and biotic DEGs were then compared to identify common genes with conserved and non-conserved gene expression i.e. whether up or down or oppositely regulated in both the categories, revealing the broad patterns of their

involvement in stress response. In order to test the efficiency of identified common DEGs in classification of abiotic and biotic stresses as well as individual stresses within abiotic and biotic stresses, we systematically investigated various classification and machine learning techniques including PCA, Partial least squares Discriminant Analysis (PLS-DA), SVM and Random Forest (RF). We characterized the shared DEGs through functional enrichment analysis of gene ontologies, metabolic pathways, transcription factor families and microRNAs targeting them. We also analyzed correlation of co-expression between the common DEGs to find sets of genes showing high co-expression and identify hub genes which show most number of edges over a very high cut-off value.

4.3 Methods

4.3.1 Selection of stress response microarray studies and identification of differentially expressed genes

All of the microarray studies performed on Affymetrix Rice Genome Array and deposited at Gene Expression Omnibus (GEO) under the platform GPL2025 were manually searched to identify and categorize 13 stress conditions (7 abiotic and 6 biotic stresses) as shown in Table S4.1. Two meta-analysis studies were performed combining abiotic and biotic stresses separately. Briefly, the raw intensity CEL files of the selected samples were downloaded from GEO and intensity values were extracted from the CEL files using the bioconductor package Affy in R (Gautier et al., 2004), quality checked using the package, ArrayQualityMetrics (Kauffmann et al., 2009) and the samples failing quality tests were removed.

The samples of each stress were normalized together using Robust Multichip Average (RMA) method (Irizarry et al., 2003). The probes were then matched to their loci based on annotation provided at ricechip.org (http://www.ricechip.org). Probes with no match or ambiguously matching multiple loci were discarded. The retained probes and their normalized intensity values were then loaded into oneChannelGUI environment to perform non-specific filtering of probes with relatively small signal distribution using

Inter Quartile Range (IQR) filter at most stringent setting (0.5) and probes with very low intensity values (probes below threshold log2(50)=5.64 in ≥90% of arrays). Differentially expressed genes (DEGs) were identified using Rank Product method (Breitling et al., 2004). We used the function RPadvance of the bioconductor package RankProd (Hong et al., 2006) which is specifically designed for meta-analysis by taking into consideration the different origins of samples. The number of permutation tests was set to 250. The function topGene with a PFP cut-off value of ≤0.01 was used to output differentially expressed genes. Among multiple probes matching the same locus, the probe ID with highest fold change was retained.

4.3.2 Classification methods

We used a number of classification and machine learning techniques to assess the performance of identified common DEGs between abiotic and biotic stresses in classification of different stresses. We extracted the RMA normalized intensity values of the identified common DEGs between abiotic and biotic stresses from stress treated microarrays (126 Abio and 232 Bio arrays) and scale adjusted using mean-centering and dividing by the square root of standard deviation of each variable (pareto scaling) (Fig. S4.1). Pareto scaling was chosen as it keeps the data structure partially intact while reducing the relative importance of large values (van den Berg et al., 2006).

Principal Component Analysis (PCA) is a non-supervised (i.e. does not make use of class labels) dimensionality reduction procedure which performs an orthogonal transformation of the original variables into a set of linearly uncorrelated variables such that the largest variance between the classes is captured in the transformed variables also called as principal components (PCs) (Yeung and Ruzzo, 2001). The PCs are numbered in decreasing order and the top PC (PC1) captures the maximal variance between different classes. Partial least squares Discriminant Analysis (PLS-DA) is a supervised (i.e. makes use of class labels) projection method that separates groups by rotating the PCs such that a maximum separation among classes is obtained (Zhang et al., 2013).

SVM classifies binary training data by drawing a hyper-plane (linear or nonlinear based on type of kernel selected) that maximally separates the two categories (Furey et

al., 2000). R-SVM performs this type of classification recursively using different feature subsets and selects the best performing features based on cross-validation error rates. Although SVM based on microarray data is widely used to classify and predict disease status in humans (Hedenfalk et al., 2001) and identify important features (Zhang et al., 2006), only a few studies have used R-SVM to identify stress responsive genes in plants (Liang et al., 2011). Random Forest (RF) is a decision tree based algorithm that grows the branches of an ensemble of classification trees by selecting random subsets of features from bootstrap samples and makes class prediction based on majority vote of the ensemble. A number of characteristics of RF make it ideal for our data set including its use for multi-class problems, less affected by noise and does not overfit the training data (Diaz-Uriarte and Alvarez de Andres, 2006). The statistical packages and tools provided by R, WEKA (Frank et al., 2004) and Metaboanalyst (Xia et al., 2012) were utilized to implement different analytical procedures.

4.3.3 Functional enrichment analysis

Gene ontology analysis was carried out using the Singular Enrichment Analysis (SEA) tool offered by agriGO (Du et al., 2010) at default settings of Fisher t-test ($p<0.05$), False Discovery Rate (FDR) correction by Hochberg method and five minimum number of mapping entries against species specific pre-computed background reference. Metabolic pathway enrichment analysis was carried out using the tool Database for Annotation, Visualization and Integrated Discovery (DAVID) v6.7 (Huang et al., 2009). Information on transcription factors (TFs) genes in rice was obtained from the database PlnTFDB (Perez-Rodriguez et al., 2010) and analyzed for enrichment of TF families. The microRNAs predicted to target stress responsive genes were obtained from plant microRNA database (Zhang et al., 2010)

4.4 Results

4.4.1 Differentially expressed genes common to abiotic and biotic stresses

We analyzed 559 microarray samples (219 from abiotic and 340 from biotic stresses) from 13 stress conditions of which 7 were abiotic (cold, drought, heat shock, metal, nutrient, salt and submergence) and 6 were biotic stresses (bacteria, fungi, insect, nematode, virus and weed) (Table S4.1A). Meta-analysis by combinatorial analysis of 7 abiotic stresses from 15 different studies together revealed 3471 differentially expressed genes (DEGs) and 6 biotic stresses from 17 different studies revealed 3065 DEGs with false discovery rate (FDR) ≤ 0.01 (Fig. 4.1A and Table S4.2). About 60% of DEGs in abiotic stresses were downregulated while 60% of DEGs in biotic stresses were upregulated (Fig. 4.1B). This broad pattern indicates that a wide variety of biological processes are downregulated under abiotic stress as it affects the whole system thus driving the plant to a protective and energy conserving mode. On the other hand, biotic stresses are often localized especially at the early stages and require an array of defense response molecules and metabolites to be synthesized and orchestrated as in for example systemic acquired resistance (SAR) to execute a resistance response against a specific infectious organism. Among the DEGs, more than 26% or 1377 genes were common to abiotic and biotic stresses indicating that these genes which are just 3.5% of all non-TE genes in rice (MSU7.0) are affected by a diverse set of stress conditions and possibly play significant roles in multiple stress responses (Table S4.3). Our major objective in this study is to analyze the stress responsive genes involved in multiple stresses that regulate cross talk between abiotic and biotic stresses. Therefore, we focused on the 1377 common DEGs for our study.

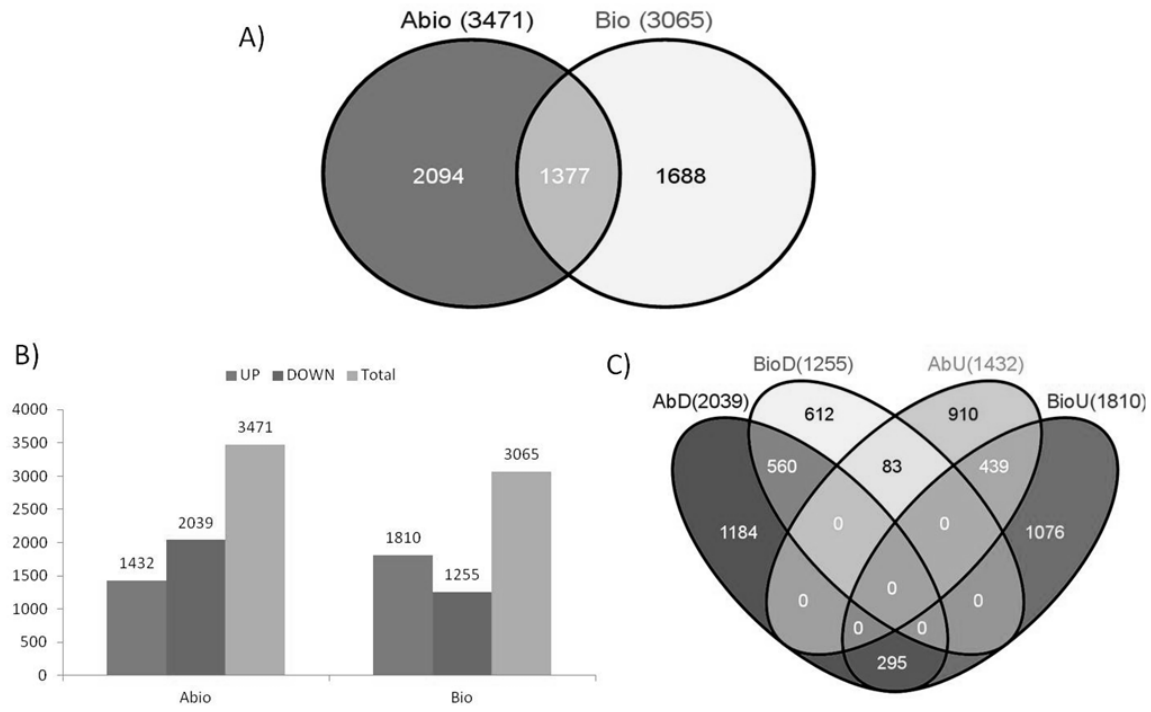

Figure 4.1: Comparison of identified differentially expressed genes in abiotic and biotic stress responses. A) Two-way venn diagrum showing the common DEGs between abiotic and biotic stresses. B) Number of up and downregulated in all of identified abiotic and biotic stresses DEGs. C) Four-way venn diagram showing number of genes showing conserved and non-conserved expression status.

We found 72% or 999 out of 1377 common DEGs with conserved expression between abiotic and biotic stresses suggesting most of these genes and their associated biological processes are regulated in a similar fashion in vast majority of stress conditions. Among the 28% of DEGs showing non-conserved expression, >21% or 295 genes were downregulated in abiotic and upregulated in biotic stress (Fig. 4.1C). In our previous study, where we compared only bacterial versus drought stress in rice, we found similar pattern with higher number of DEGs downregulated in drought and higher number of DEGs upregulated under bacterial stress among the DEGs common to these two stresses. About 16% or 221 of these genes are annotated as 'expressed protein' and ~7% or 96 have no GOSlim assignment revealing that many of stress responsive genes are still poorly understood. Studies elucidating functional roles of these genes would be vital for comprehensive understanding of stress response in rice.

4.4.2 Machine learning approaches based on common DEGs classified abiotic and biotic stresses into two classes with high accuracy

Using the identified common DEGs, we investigated if the different stress conditions can be accurately classified using machine learning approaches. Initially, we investigated the performance of PCA in discriminating abiotic stresses from biotic stresses as two-classes using all of the 1377 common DEGs. The first three PCs captured 56.4% of variance between the samples. The 3D-PCA plot of top 3 PCs showed clear separation of abiotic and biotic classes for majority of the samples although both the classes were widely dispersed across components (Fig. 4.2A). Nonetheless, there were some samples showing considerable overlap between the classes. We then analyzed the data-set using Partial least squares Discriminant Analysis (PLS-DA), a technique that is specifically suited for analysis of data-set with high feature dimensions and multicollinearity (Perez-Enciso and Tenenhaus, 2003). Many of the published microarray studies have found PLS-DA as a highly efficient method for multiclass classification (Student and Fujarewicz, 2012). PLS-DA resulted in five components which captured ~62% variance between the two classes and separated them with a very high accuracy of 0.99 (R^2:0.95 (goodness of fit), Q^2: 0.93 (predictive value) p-val <0.01) upon 10 fold cross-validation. The 3D plot of PLS-DA showed clear separation of all of the samples between abiotic and biotic stresses (Fig. 4.2B). The important genes contributing most to the PLS-DA separation can be identified using Variable Importance in Projection (VIP) score which is a weighted sum of squares of PLS loadings (Perez-Enciso and Tenenhaus, 2003). There were 177 genes with the VIP-score (component 1) cutoff value ≥1.5 (Zhang et al., 2013) and 33 genes with values ≥2 (Table S4.4).

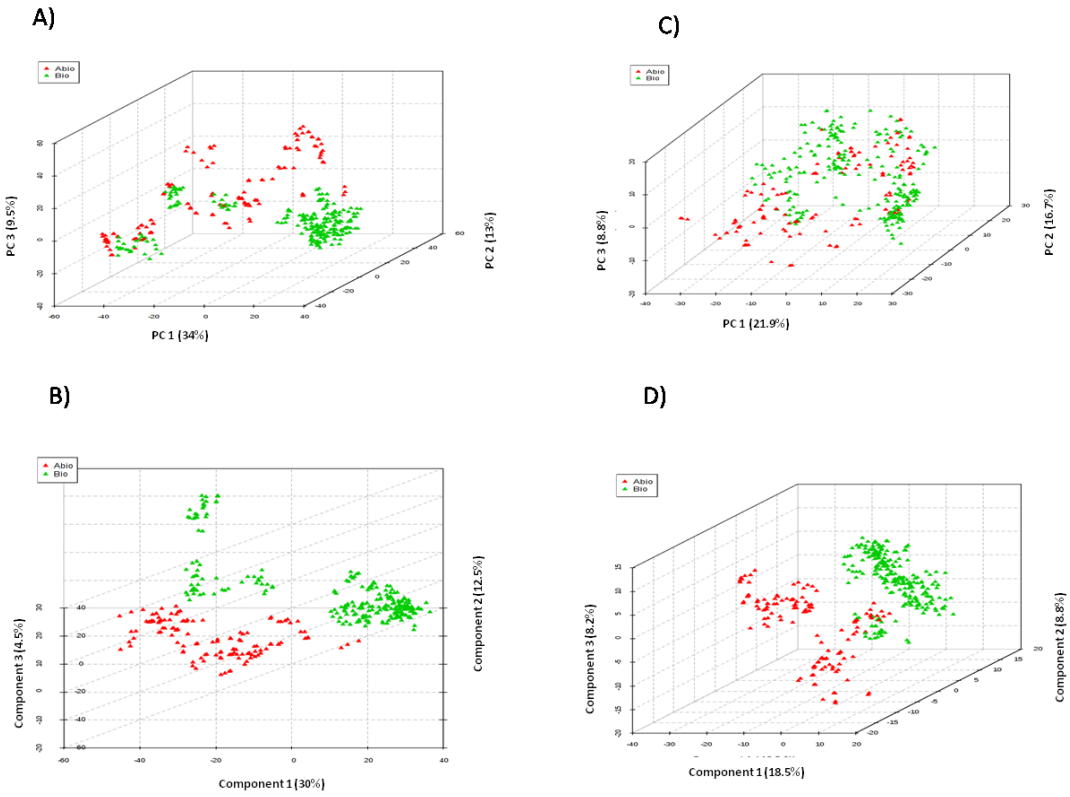

Figure 4.2: Three dimensional plots of two-class Classification of abiotic and biotic stresses. (A) and (B) show 3D plots based on top three components by PCA and PLS-DA, respectively using all of 1377 common DEGs. (C) and (D) show 3D plots based on top three components by PCA and PLS-DA, respectively using top 540 genes ranked by SVM to return 100% accuracy of classification.

Next, we analyzed the same data-set using another very popular supervised learning technique for microarray data classification called Recursive-Support Vector Machine (R-SVM) which identified 540 genes (39.2% out of 1377) that can classify abiotic and biotic stresses with 100% accuracy and 88 (6%) genes with 95% accuracy (Fig. 4.3). These 540 genes included a number of hormone response and stress response signaling genes. All five of the MYB TFs which are important regulators of development and defense responses in plants (Yanhui et al., 2006) found in the common DEGs were part of these 540 genes. Further, 103 (19%) of the 540 genes were part of a recently published database namely stress-responsive transcription factor database (STIFDB2) (Naika et al., 2013) which provides a list of stress responsive genes (1118 genes of *Oryza sativa* subsp. *japonica*) identified through biocuration and genomic data mining. Out of

540 genes, 178 (33%) were the ones with non-conserved expression pattern between abiotic and biotic stresses, which is slightly higher compared to the 28% of genes showing non-conserved expression in all of the common DEGs. Although PCA based on these 540 genes resulted in poor separation of the classes with 47.4% variance captured by top 3 PCs, PLS-DA showed clear separation of the two classes (Fig. 4.2C and D). The top 5 components of PLS-DA captured 53% of variance with classification accuracy of 0.97 (R^2:0.91, Q^2: 0.87 p-val <0.01) which is slightly less than the 0.99 accuracy obtained using all of the 1377 common DEGs. There were 79 genes (14% of 540) with the VIP-scores ≥1.5 and 27 genes with ≥ 2. There were two genes with VIP-scores ≥3 which code for xylanase inhibitor and glycosyl hydrolase both showing conserved upregulation.

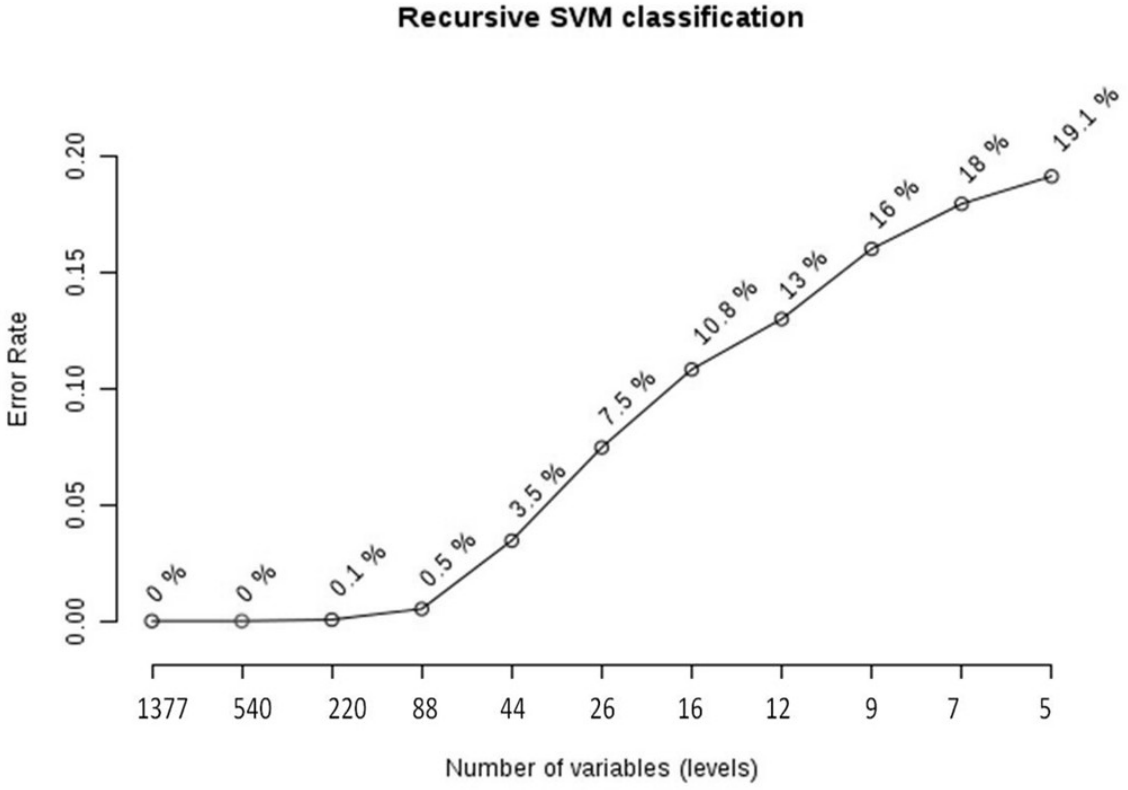

Figure 4.3: Classification error rates of different subsets of common DEGs upon 10-fold Cross Validation (CV) using R-SVM. Error rate using all of 1377 or 540 common DEGs was 0% (or 100% accuracy of classification) and 0.1% (99% of accuracy) using 220 genes and 0.5% (95% accuracy using 88 genes)

4.4.3 Analysis of common DEGs identified top genes with discordant behavior among multiple stresses

From the 13 stress conditions analyzed, we selected top 10 stresses (5 abiotic stresses: drought, metal, salt, cold and nutrient and 5 biotic stresses: bacteria, fungus, insect, weed and nematode) based on higher number of microarray samples. We analyzed this data using the normalized and pareto scaled intensities of 1377 DEGs to assess the performance of these genes in classification of different stress conditions. The top five components of PLS-DA captured 62.9% of variance between various stresses and showed classification accuracy of 0.77 (R^2:0.92, Q^2: 0.88 p-val <0.01). There were 196 and 53 genes with VIP scores (component 1) ≥1.5 and ≥2. The relatively low classification accuracy reflects the inherent similar expression patterns between different stresses. Nonetheless, the components 1 and 3 as shown in the 2D score plot and top three components as shown in 3D score plot were able to clearly separate abiotic and biotic stresses as two major groups (Fig. 4.4). The 2D and 3D plots also showed wide dispersion of drought stress and closeness with majority of cold stress samples. Similarly, the 3D plot showed higher overlap between salt and metal stresses than other stresses suggesting higher similarity of gene expression profile between them. The nutrient stress samples can be observed as a distinct group although closer to other abiotic stresses. Bacterial stress samples show two major groups. One of the groups with most of bacterial samples showed overlap with fungal stress samples only. The other group was closer to weed, nematode and fungal stress samples. Insect stress was observed as a distinct group closer to the group with bacterial and fungal samples.

The same data-set was analyzed using another classification technique called Random Forest (RF) which classified 8 of the 10 stresses with 100% accuracy with an overall out-of-box (OOB) error rate of 0.0087 which is an unbiased estimate of classification error based on the one third left out samples (test samples) after bootstrap sample selection (Table 4.1). Two of the stresses with less than 100% accuracy of classification were salt with one wrongly classified sample (error rate: 0.037) and fungal stress with two wrongly classified samples (error rate: 0.08). RF also provides a measure of variable importance by evaluating the increase in OOB error rate upon permutations

called mean decrease in accuracy (Hsueh et al., 2013). The top 15 significant genes based on mean decrease in accuracy are shown in figure S2 including, LOC_Os02g45170 (error rate: 0.0056) a bHLH TF and LOC_Os05g31040, which codes for cytokinin dehydrogenase precursor.

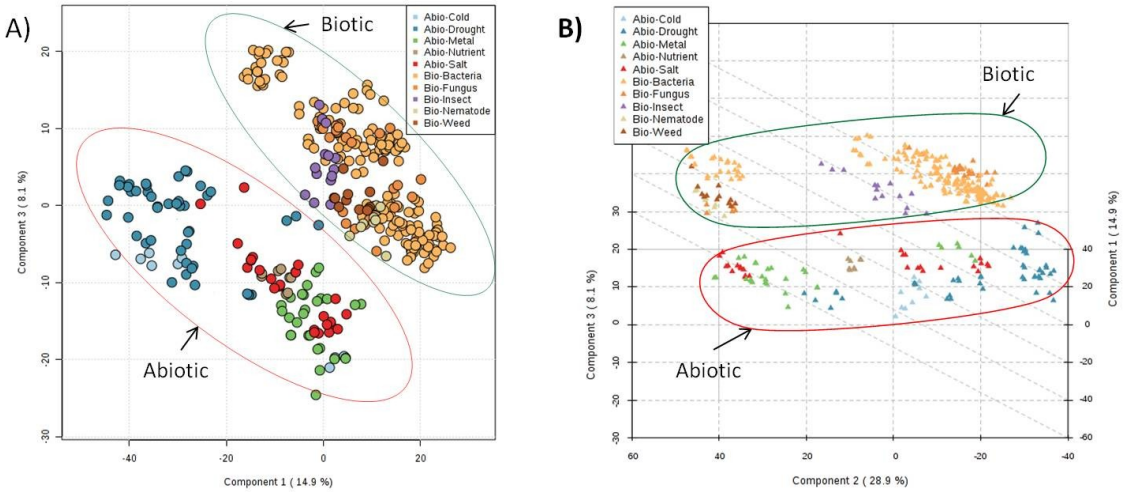

Figure 4.4: Multi-class classification of ten stress conditions by PLS-DA. All five abiotic stresses are circled by a red oval and all five biotic stresses by a green oval. A) Two-D plot between PLS-DA components 1 (14.9%) and 3 (8.1). B) Three-D plot between PLS-DA componenets 1 (14.9%), 2 (28.9%) and 3 (8.1%)

4.4.4 Functional enrichment analysis revealed enrichment of distinct molecular mechanisms and gene families in conserved and non-conserved gene sets

Gene ontology enrichment analysis of the 560 genes showing conserved downregulation in abiotic and biotic stresses revealed enrichment of many of the major biological and cellular processes including photosynthesis (FDR: 1.40E-07), electron carrier activity (FDR: 3.60E-06), small molecule biosynthetic process (FDR: 2.10E-05), cellular nitrogen compound metabolic process which is the parent term for a number of amino acids and nucleobase-containing compounds. The terms transcription repressor activity (FDR: 0.0008) and response to oxidative stress (FDR: 0.034) were also found to be significant (Fig. S4.3 and table S4.4). On the other hand, 439 genes showing conserved upregulation revealed a number of terms related to regulatory processes. The most significant innermost child terms are serine-type endopeptidase inhibitor activity (FDR: 2.2E-06),

chitin catabolic process (FDR: 0.00013), cell wall macromolecule catabolic process and regulation of transcription. Serine proteases serve diverse set of physiological roles in plants, important among which are induction after pathogen attack leading to hypersensitivity response (HR), regulation of Rubisco proteolysis, stomata development, perception of growth hormones, symbiosis and senescence (Antao and Malcata, 2005, van der Hoorn, 2008). Significant enrichment of inhibitors of serine-type endopeptidases in diverse stress conditions indicates induction of many of the activities repressed by serine proteases as part of stress response. Further, serine protease inhibitors were also found to act as defense proteins by suppressing the activity of bowel proteinases in insects and plant pathogenic microorganisms (Mosolov and Valueva, 2011). Among the genes showing non-conserved expression, the set of genes downregulated in abiotic stresses and upregulated in biotic stresses were enriched with GO terms, extracellular region (FDR: 5.30E-06), catalytic activity (5.3E-05), reproduction (FDR: 0.0041), kinase activity, response to stress and transcription factor activity.

The functional annotation tool DAVID (Database for Annotation, Visualization and Integrated Discovery) v6.7 is an excellent tool that performs enrichment analysis of various annotation resources including gene ontologies, protein domains and pathways using a modified Fisher exact test called EASE. Further, it clusters significant annotation terms using kappa statistics and fuzzy heuristic clustering based on the degree of common genes between two annotations and provides an enrichment score for each annotation cluster. In the conserved downregulated gene set, there were four annotation clusters with enrichment score >2.0 related to porphyrin and chlorophyll metabolism, transcription repressor activity via Nmr-A like domain which is involved in post-translational modification of the GATA-transcription factors (Stammers et al., 2001), photosynthesis and nicotianamine synthase activity. There were three annotation clusters with enrichment score >2.0 in conserved upregulated gene set related to Heat shock protein Hsp20, valine, leucine and isoleucine degradation, and Bowman-Birk proteinase inhibitor (BBPI) family of serine protease inhibitors. In rice, BBPI genes were reported previously to be induced in multiple stresses like wounding, infection and hormonal stress (Rakwal et al., 2001, Qu et al., 2003). The top annotation clusters in the non-

conserved abiotic down and biotic up gene set were made up of a number of interpro domain terms, glycoprotein, metal-ion binding, plant peroxidases and glycoside hydrolases.

There were 97 transcription factor and regulator genes in the common DEGs (7%) belonging to 24 gene families. A distinct pattern of overlap between conserved downregulated genes and non-conserved abio down and bio up gene sets in the major TF families NAC, HSF, WRKY, MYB, MYB_related and C2H2, and conserved upregulated genes and non-conserved abio up and bio down gene sets in the major TF families ERF, bZIP and C2H2 was observed (Table S4.5). Twelve out of thirteen Ethylene Response Factors (ERFs) were found in conserved upregulated gene sets. These AP2 (APETALA2) domain containing ERFs are well known for their role in both abiotic and biotic stress responses and were also shown to enhance multiple stress tolerance (Xu et al., 2011). Nine out of twelve WRKY TFs were part of non-conserved abiotic down and biotic up gene set which suggests that these TFs (WRKY24, 28, 45, 47, 62, 71, 72, 76 and 79) respond differently to abiotic and biotic stress signals and are the major regulatory factors that determine the direction of molecular machinery and ultimately the cellular fate under simultaneous multiple stresses. All of the thirteen MYB and MYB_related TFs in the common DEGs were downregulated in abiotic stresses while five of them were upregulated in biotic stresses. MYB along with NAC TFs are reported to control antagonism between hormone-mediated abiotic stress and pathogen response pathways (Atkinson and Urwin, 2012). On the other hand, all five of G2 (Golden2)-like TF family members which also contain MYB-like DNA binding domain were part of conserved upregulated gene set. The G2-like TFs are required for proper chloroplast development and were shown to influence nuclear photosynthetic gene expression (Waters et al., 2009). We found a dearth of studies on the role of G2-like TFs under stress conditions. Downregulation of photosynthetic mechanisms under stress is well established as also observed in the enriched GO terms in our conserved downregulated gene set. Careful manipulation of G2-like TFs would shed further light on regulation of photosynthesis under stress and reveal novel mechanisms to enhance stress tolerance. Out of the five LSD (Lesion Simulating Disease) (Dietrich et al., 1997) family members reported in

Oryza sativa subsp. *japonica* by PlnTFDB, two were part of conserved downregulated gene set. LSD TFs act as negative regulators of programmed cell death (PCD) in a hypersensitive response (HR) (Epple et al., 2003). Transgenic suppression of LSD orthologs in rice resulted in a dwarf phenotype due to deficiency of bioactive gibberllin while overexpression of LSD enhanced resistance to rice bacterial blight (Xu and He, 2007). Based on our finding, studying LSD TFs under simultaneous abiotic and biotic stresses would provide vital clues on stress cross-talk and modulation of PCD.

We analyzed the microRNAs predicted to target the 1377 common DEGs using the database PMRD (Zhang et al., 2010). Out of the 456 experimentally verified miRNAs (miRBase (Griffiths-Jones et al., 2006)) in rice, 142 (31%) miRNAs belonging to 50 miRNA families were found to target one or more common DEGs (Table S4.6). Recently, 35 miRNAs from 31 miRNA families were found to be differentially expressed under abiotic stresses, drought, salt and cold (Shen et al., 2010). Eighteen of these 31 stress responsive miRNA families were part of the 50 miRNA families targeting the common DEGs. The miRNA osa-miR1436 was found to target five of the conserved upregulated genes including LOC_Os09g23620, a MYB TF while osa-miR446 was found to target five of the conserved downregulated genes.

4.4.5 Co-expression analysis revealed two dense clusters of positively and negatively correlated genes under multiple stresses

We conducted co-expression analysis using the normalized gene expression values of the common DEGs from stressed microarray samples and calculating Pearson Correlation Coefficient (r) between them. Out of the 947,376 possible edges (co-expression gene pairs) between the common DEGs, we found 8,924 edges with very high correlation ($r \geq 0.9$ = 4,254 and $r \geq -0.7$ = 4,670 edges, p-value = 0.01) in abiotic stress samples and 21,229 edges ($r \geq 0.9$ = 7,673 and $r \geq -0.7$ = 13,656 edges, p-value = 0.01) in biotic stress samples. A very high number of negative edges were observed in biotic stresses compared to abiotic stresses. For instance, there were 88 edges in biotic stresses with $r \geq -0.9$ but only four edges in abiotic stresses with $r \geq -0.9$. There were 3,701 shared edges between the two data-sets with $r \geq 0.9$ and $r \geq -0.7$, out of which 2,684 (72%) were

positive edges and 1,017 were negative edges. These 3,701 edges were between 381 genes, out of which 257 (67%) genes showed conserved downregulation, 54 genes showed conserved upregulation and 49 genes showed downregulation in abiotic stresses and upregulation in biotic stresses. The 2,684 positive edges were between 208 genes, out of which 194 (93%) were the genes that showed conserved downregulation. Among the 381 genes, 15 had >75 high correlation edges. The top three genes with most number of edges were, LOC_Os02g22480 (glycosyltransferase -142 edges), LOC_Os11g47840 (Putative Rhomboid homologue - 120) and LOC_Os03g57200 (glutathione S-transferase - 93). All three of these genes showed conserved upregulation. Among the 14 TFs with significant edges, three TFs belonging to NF-Y (Nuclear Factor –Y, a histone like CCAAT-binding domain TF), G2-like and bHLH TF families had most number of significant edges (79, 37 and 20, respectively). Majority of these edges were positive edges with other genes that showed conserved downregulation.

We analyzed the 3,701 significant edges using the plugin NetworkAnalyzer in network analysis platform Cytocape 2.8.3 (Shannon et al., 2003) which revealed a dense cluster of positive edges (edges with $r \geq 0.95$ are shown in red color) which included most of the nodes with >75 edges (shown in blue) and a sparse cluster of negative edges (edges with $r \geq -0.9$ are shown in green) (Fig. 4.5). The two positive edge and negative edge rich clusters were found to be bridged by the gene LOC_Os01g13570, coding for phosphoglycerate mutase with a positive edge to SOUL heme-binding protein that was highly connected to negative edge rich cluster and positive edges with rhodanese and pentatricopeptide (PPR) domain containing proteins which were highly connected to the positive edge rich cluster.

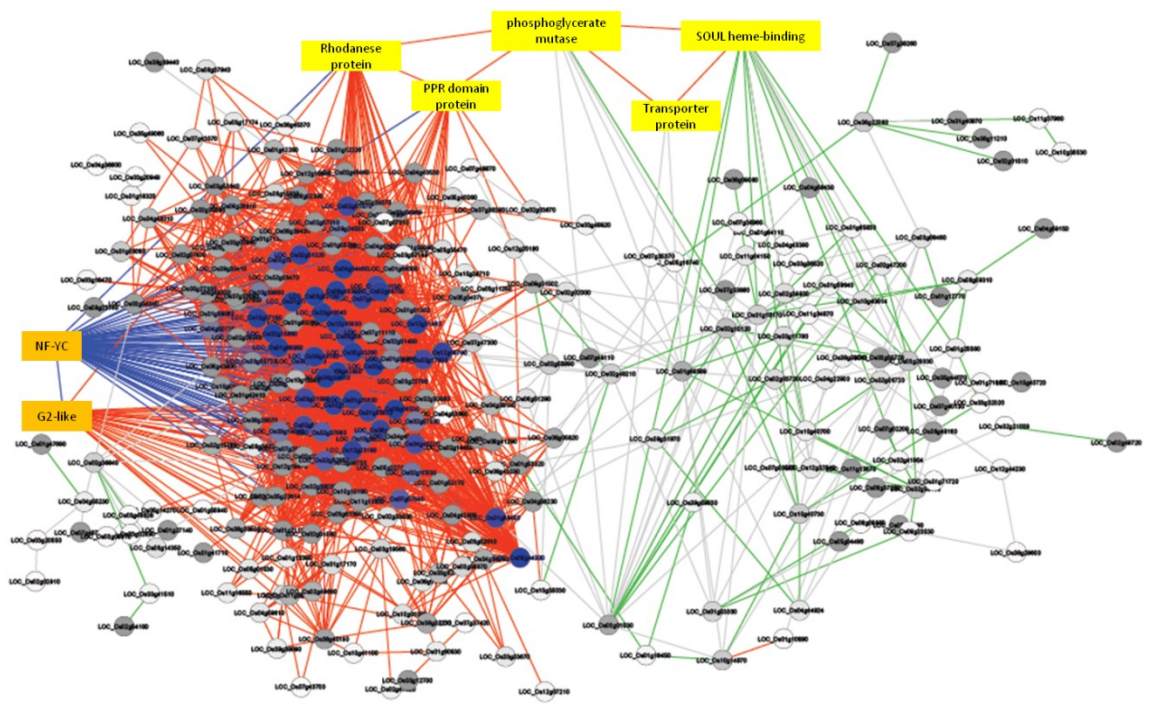

Figure 4.5: Co-expression network of common DEGs. The edges with r ≥ 0.95 are shown in red and r ≥ -0.9 are show in green. Nodes with >75 edges are shown in blue and >25 are shown in grey. The edges of NF-YC TF are shown in blue.

4.4.6 High overlap between top genes identified by different classification techniques, co-expression and functional enrichment analysis

We compiled the significance of the common DEGs based on various criteria including feature importance as found by different classification techniques, count of number of co-expression edges, PlnTFDB gene, and STIFDB2 gene (Table S4.3). We found that many of the PLS-DA two-class significant genes (177 genes with VIP ≥ 1.5) were also significant in PLS-DA multiclass (36% or 71 out 196) and RF's top 100 genes (68%) but showed poor overlap with the 540 significant genes found by SVM (2% or 9 out of 540), TF genes (9% or 9 out of 97) and STIFDB2 genes (10% or 27 out of 259). However SVM's 540 genes showed high overlap with PLS-DA multiclass (50% or 99 out of 196), TF genes (45% or 44 out of 97) and STIFDB2 genes (40% or 103 out of 259). Taken together, the 196 top genes of PLS-DA multiclass showed overlap with most of the other significant gene lists, of which 43 (22%) were also part of STIFDB2 list. Out

of 1,118 *Oryza sativa* subsp. *japonica* genes reported as stress responsive genes in STIFDB2 (Naika et al., 2013), 259 (23%) were part of common DEGs. Further, out of 97 TF genes in the common DEGs, only 12 were part of STIFDB2 and none of the major WRKY and MYB TF genes including those previously reported as stress responsive genes (Atkinson and Urwin, 2012) were part of STIFDB2's list, which indicates that it is not a comprehensive database for rice. The top 10 of these 196 genes are given in table 4.2. The topmost gene encodes a CCCH zinc finger domain containing TF known to control embryogenesis (Li and Thomas, 1998) and involved in multiple abiotic stresses (Sun et al., 2007, Kim et al., 2008). A homolog of this gene (LOC_Os05g10670) which was also part of the 1432 upregulated genes in our meta-analysis of abiotic stresses, was recently reported to confer delayed senescence and improved tolerance to high-salt and drought stresses by regulating reactive oxygen species homeostasis, and metal homeostasis (Jan et al., 2013). One gene which was part of all feature selection lists was LOC_Os11g26780, a dehydrin gene which had one significant positive edge with another dehydrin gene (LOC_Os11g26790, r=0.97 and 0.93 in abiotic and biotic stresses, respectively) both of which showed conserved upregulation.

Comparison of the common DEGs with the list of 1922 hormone related genes of *Arabidopsis* as reported in Arabidopsis Hormone Database 2.0 (Jiang et al., 2011) using putative orthologous genes found by GreenPhylDB (Rouard et al., 2011) revealed 31 common DEGs that were orthologous to 51 *Arabidopsis* hormone genes (Table S4.3). A summary table of the expression status of hormone related genes in the common DEGs (78 genes) based on Arabidopsis hormone database orthologs and paralogs with same annotation and expression status in both abiotic and biotic stresses (except TFs) or name of the hormone in the gene annotation provided by MSU 7.1 is given in Table 4.3. Overall, the distribution of expression status of various hormone related genes was very similar to the one proposed in a recent review (Atkinson and Urwin, 2012). For instance, 9 out 12 abscisic acid responsive genes showed conserved upregulation while 6 out of 10 ethylene responsive genes showed non-conserved abiotic down and biotic upregulation. Most of the conserved downregulated genes of auxin were related to auxin biosynthesis and response factors while conserved upregulated were related to auxin repressed factors

which indicates extensive downregulation of auxin induced biological processes. A recent study analyzed transcriptome of rice under bacterial stress by *Xanthomonas oryzae pv. oryzae* and compared the DEGs with those found in seven other microarray studies conducted on Affymetrix RiceArray (Narsai et al., 2013). They reported 240 genes (212 loci) as differentially expressed in multiple stresses. Out of these loci, 110 (51.8%) were part of our common DEGs list, most of which belonged to conserved upregulation geneset (64%) and included many important genes such as WRKY, AP2/EREBP TFs, ABC transporter, multidrug resistance and universal stress genes.

4.5 Discussion

Multiple stress response in plants has been a hot topic of research as many studies, including those involving genetic manipulation and chemical intervention reported increased resistance to one stress resulted in heightened susceptibility to other abiotic and biotic stress conditions (Atkinson and Urwin, 2012, Sharma et al., 2013). Further, it was suggested that plant hormones are the key determinates of genetic switches and cellular adjustments in a multi-stress environment. Different plant hormones are broadly categorized to play central roles in different kinds of stress responses. For instance, within biotic stresses, (hemi)biotrophic pathogens commonly activate salicylic acid (SA)-dependent defense response, while necrotrophic pathogens activate jasmonic acid (JA) and ethylene (ET)-dependent signaling pathways (Sharma et al., 2013). SA and JA/ET often act antagonistically and propagate opposing influences (Pieterse et al., 2009). On the other hand, abscisic acid (ABA) is well established as the major player of abiotic stress response. ABA is increasingly found to also play a critical role in biotic stresses by negatively regulating plant immunity. Many studies found that abiotic stresses enhance plant susceptibility to pathogen attacks due to weakening of defense systems. Thus, it was proposed that plants prioritize abiotic stress tolerance over biotic stress response with ABA as molecular switch between the two responses to minimize the damage (Lee and Luan, 2012). Recently, however, contrary studies where biotic stress takes precedence are also reported (Kim et al., 2011, Mang et al., 2012, Sanchez-Vallet et al., 2012). Thus, in light of these recent developments which revealed a rather complicated picture of

multiple stress response, we embarked on identification of differentially expressed genes in abiotic and biotic stress environments separately and performed comparative analysis of the shared stress responsive genes, which would provide vital clues on the causative factors behind the cross-talk resulting in the observed synergistic and antagonistic regulation of known abiotic and biotic stress response pathways.

We conducted meta-analysis of publically available microarray studies on a diverse set of stresses in rice from the same microarray platform and found about 5159 DEGs (3471 genes under abiotic stresses and 3065 biotic stresses) which represent an exhaustive list of genes involved in stress response in rice. Although, we utilized a single microarray platrfrom and robust statistical methods including QC by ArrayQualityMetrics to filter out samples failing quality tests and oneChannelGUI to filter out probes with very low expression values or IQR, RMA normalization, RPadvance which is a differential expression detection method specifically designed for meta-analysis that takes into consideration different origins of samples and stringent cut-off value (FDR $\leq$0.01) we cannot rule out heterogeneity caused due to various factors like differences in the basal expression level or stress tolerance in between different cultivars or ecotypes (Table S4.1). Thus, such factors should be taken in consideration in the interpretation and application of the findings in this study.

Among the 5159 DEGs found in both types of stress conditions, there were 1377 (26.6%) common genes. As these genes were found by combinatorial analysis of a wide spectrum of abiotic and biotic stress conditions, their expression status can be considered as a representation of their overall involvement in stress response to non-living factors and living organisms. Thus, this list of genes forms an ideal geneset to objectively investigate the similarities and differences between abiotic and biotic stress responses. Although >70% of common DEGs showed conserved differential expression, we were able to classify different stresses (including abiotic and biotic stresses) with high accuracy indicating the subtle expression differences of these genes can be exploited to effectively discriminate between various stress conditions.

A closer look at chloroplast and photosynthesis related genes in the common DEGs revealed conserved downregulation of 17 out of 18 phtotosystem II, chlorophyll

A-B binding and thylakoid lumenal genes (except LOC_Os04g59440 which showed non-conserved upregulation in abiotic stress) (Fig. S4.4). A diverse set of 40 chloroplast precursor enzymes which contain an amino-terminal transit peptide for import into chloroplast (Jarvis, 2008) were also part of the common DEGs, 26 (65%) of which showed conserved downregulation. Further, a number of cytochrome P450 genes (29 genes) which are parent compounds for a number of secondary metabolites involved in plant defense (Jirschitzka et al., 2013) were part of the common DEGs. Fourteen of these 29 (~48%) genes showed conserved downregulation, while 7 showed conserved upregulation and the rest showed non-conserved differential expression indicating an important role for these genes in abiotic stress response. Thus, exploring the non-conserved DEGs would shed further light on the cross-talk of stress response via metabolic adjustments. Cell wall is the first line of plant defense in response to external stimuli. A number of important gene families involved in cell wall synthesis and modifications showed distinct patterns of expression under abiotic and biotic stresses. For instance, there were 6 OsWAK (Wall Associated Kinase) genes in common DEGs, all of which showed non-conserved downregulation under abiotic stresses and upregulation under biotic stresses. WAKs are part of the transmembrane Receptor-Like Kinase (RLK) superfamily which perceive stimuli by their extracellular domains and transmit the signals via their cytoplasmic kinase domains (Li et al., 2009). There are currently 144 genes regarded as WAKs (MSU7.0) compared to 26 genes in *Arabidopsis* which is most likely due to lineage specific gene duplications (Zhang et al., 2005). However, very little is known about the function of most of these genes in rice except OsWAK1, which was found to increase resistance to blast fungus, *Magnaporthe oryzae* upon overexpression (Li et al., 2009, Kohorn and Kohorn, 2012). FAS1 (fasciclin-like) domain containing genes are another group of transmembrane genes involved in cell adhesion (Johnson et al., 2003, Ma and Zhao, 2010). All 5 of fasciclin domain genes in the common DEGs showed conserved downregulation. Similarly, most of cupin, expansin and aquaporin genes involved in cell wall synthesis and organization showed conserved downregulation. Further, chitinase and laccase genes were highly downregulated especially in biotic

stresses indicating cell wall reorganization as an integral part of plant defense system against a wide range of pathogens.

A number of transporter genes showed clear patterns of difference in expression between abiotic and biotic stresses. For instance, 2 out of 3 major facilitator superfamily (MFS) antiporter genes showed non-conserved upregulation under abiotic stresses. All three of the genes coding for pleiotropic drug resistance (PDR) type ATP-binding cassette (ABC) transporter proteins which were found to be induced by ABA, SA and Jasmonate in rice (Moons, 2008) showed conserved upregulation. Reversible protein phosphorylation executed by kinases and phosphatases is a fundamental mechanism that facilitates the orchestration of some of the most sophisticated signaling pathways. A number of different kinds of kinases and phosphatases were found in the list of common DEGs out of which Ser/Thr protein kinases and phosphatases showed high distinction between the two stresses as also found by the GO analysis (Table S4.4). All five of the protein phosphatase 2C (PP2C) genes showed conserved upregulation which are key players in ABA signaling pathways. Four of these PP2C genes were part of the significant genes found by both SVM and PLS-DA multi-class indicating that these genes show distinct pattern of expression in different stress conditions and can be considered as some of the most important genes to study multiple stress response. As many as 23 peroxidase (POX) precursor genes were part of common DEGs, out of which 13 (56%) showed non-conserved downregulation under abiotic stresses. Further, 9 and 12 of these 23 POX genes were part of SVM and PLS-DA multi-class significant features, respectively. A study on rice infected with blast fungus showed ten POX genes redundantly respond to multiple stresses (Sasaki et al., 2004). Our findings suggest that the functionalities of many of the POX genes are specific to biotic stresses and are promising candidates to decipher the cross-talk between stresses.

The domain family with most number of conserved upregulated genes was Zinc Finger (ZF) family (including C2H2, C3H TFs, C3HC4 and ZIM domain containing members) with 14 and 15 members out of 17 showing overexpression in abiotic and biotic stresses, respectively. All of the eleven pentatricopeptide (PPR) domain genes which play essential roles in RNA editing, organelle biogenesis (Yuan and Liu, 2012)

and plant development by coordinating interaction between mitochondria and chloroplasts (Toda et al., 2012) showed conserved downregulation except LOC_Os07g36450 which showed conserved upregulation. Thus, this gene would be an important candidate to further explore and understand their specific role under stress conditions and determine what makes it different from other PPR genes. Another interesting gene family showing high distinction between the two stress categories was LTP (protease inhibitor/seed storage/lipid transfer protein) with 5 out of 9 members showing non-conserved downregulation in abiotic stresses. VQ domain containing proteins were recently found to interact with WRKY TFs (WRKY33) in *Arabidopsis*. Further, knockout or overexpression of VQ substantially altered defense response (Cheng et al., 2012). There are 5 VQ domain genes in common DEGs out of which 4 showed non-conserved biotic upregulation. Further, WRKY24 which is the rice ortholog of WRKY33 also showed non-conserved biotic upregulation. The striking contrast of these set of genes in their behavior between abiotic and biotic stresses suggests them as important candidates to explore multiple stress response.

A list of studies that over-expressed or suppressed ten of the common DEGs that significantly altered the stress response are provided in table 4.4. Seven of these are TF genes and are part of significant features found by SVM. LOC_Os07g40290 is an auxin responsive gene showing conserved upregulation which also co-expressed with 40 other common DEGs. Further, we compared the common DEGs against a recently released database of *Arabidopsis* loss-of-function mutants (Lloyd and Meinke, 2012) using orthologs IDs which revealed 138 orthologous mutant genes out of which 33 showed increased resistance or sensitivity to a variety of stresses (Table S4.7).

Our observations such as high overlap with the lists of multiple stress response genes reported in STIFDB2 (Naika et al., 2013), agreement with the proposed role of hormone response genes in the stress cross-talk (Atkinson and Urwin, 2012) (Table S4.3), enrichment of a number of biological processes such as conserved downregulation of photosynthesis, electron carrier activity and nitrogen metabolism and, conserved upregulation of cell wall and chitin catabolism, regulation of transcription and serine-type endopeptidase inhibitor activity (Fig. S4.3 and Table S4.4) similar to the findings of a

number of individual abiotic and biotic stress studies (Rabbani et al., 2003, Ribot et al., 2008, Chaves et al., 2009, Lodha and Basak, 2012, Narsai et al., 2013) provide additional evidence for the utility of the common DEGs in discriminating abiotic and biotic stress responses in rice. Further, accurate two-class and multi-class classification of multiple stress conditions using different classification techniques and a portion of genes found as the top contributors to the classification, indicates this list of top genes are of high priority to understand simultaneous multiple stress response.

Table 4.1. Classification of multiple stresses using Random Forest method

	Abio-Cold	Abio-Drought	Abio-Metal	Abio-Nutrient	Abio-Salt	Bio-Bacteria	Bio-Fungus	Bio-Insect	Bio-Nematode	Bio-Weed	Class error
Abio-Cold	9	0	0	0	0	0	0	0	0	0	0
Abio-Drought	0	46	0	0	0	0	0	0	0	0	0
Abio-Metal	0	0	29	0	0	0	0	0	0	0	0
Abio-Nutrient	0	0	0	8	0	0	0	0	0	0	0
Abio-Salt	0	1	0	0	26	0	0	0	0	0	0.037
Bio-Bacteria	0	0	0	0	0	166	0	0	0	0	0
Bio-Fungus	0	0	0	0	0	2	23	0	0	0	0.08
Bio-Insect	0	0	0	0	0	0	0	14	0	0	0
Bio-Nematode	0	0	0	0	0	0	0	0	9	0	0
Bio-Weed	0	0	0	0	0	0	0	0	0	12	0
The overall Out-Of-Box (OOB) error rate was 0.0087											

Table 4.2. Top 10 genes with highest VIP (Variable Importance in Projection) score in multi-class classification by PLS-DA

MSU ID	Annotation	PLS-DA Multiple Stress (VIP comp.1)	PLS-DA AbioVsBio (VIP comp.1)	RF top 100 (Mean Decrease Accuracy)	PLS-DA two-class (on SVM 540)	SVM Sig.540 (Freq)
LOC_Os01g09620	zinc finger/CCCH transcription factor	2.899	2.1512	0	0	0
LOC_Os11g11970	expressed protein	2.8141	2.1813	0.003567	0	0
LOC_Os11g26780	Dehydrin	2.8021	1.6348	0.002054	2.5939	358
LOC_Os07g48020	peroxidase	2.7992	1.8946	0	0	0
LOC_Os06g24990	xylanase inhibitor protein 1	2.7427	2.2645	0	3.593	358
LOC_Os11g32890	expressed protein	2.718	1.6966	0	2.6919	358
LOC_Os06g48300	protein phosphatase 2C	2.6559	0	0	2.2334	358
LOC_Os10g40040	expressed protein	2.5979	0	0	2.3424	358
LOC_Os09g07350	fasciclin-like arabinogalactan protein 8	2.5059	0	0	1.9348	358
LOC_Os05g06920	relA-SpoT like protein RSH4	2.5036	2.8245	0.004242	0	0

Table 4.3. Distribution of number of expression status of various hormone related genes in the common DEGs

Hormone	Total genes	Cons. Down	Cons. Up	Non-cons (AbioUp-BioDown)	Non-cons (AbioDown-BioUp)
abscisic acid	12 (9)	-	9 (3)	-	3 (1)
Auxin	21 (6)	5 (1)	11 (3)	3 (1)	2 (1)
brassinosteroid	11 (5)	-	7 (4)	-	4 (1)
Cytokinin	4 (2)	4 (2)	-	-	-
Ethylene	10 (5)	3 (2)	1 (1)	-	6 (2)
Gibberellins	8 (3)	2 (1)	5 (2)	1 (0)	-
Jasmonic acid	7 (3)	4 (1)	-	1 (1)	2 (1)
salicylic acid	5 (3)	-	4 (2)	-	1 (1)

*- Number of orthologs of Arabidopsis plant hormone database genes are shown in brackets

Table 4.4: List of common DEGs which showed alteration in stress response upon over-expression/suppression

MSU ID	Annotation	Phenotype	Reference
LOC_Os01g55940	OsGH3.2 - Probable indole-3-acetic acid-amido synthetase, expressed	Enhanced broad spectrum disease resistance	(Fu et al., 2011)
LOC_Os02g08440	WRKY71, expressed	Enhanced defense response	(Liu et al., 2007)
LOC_Os03g60080	NAC domain-containing protein 67, putative, expressed	Increased drought and salt tolerance	(Hu et al., 2006)
LOC_Os05g25770	WRKY45, expressed	Increased susceptibility to bacteria	(Tao et al., 2009)
LOC_Os06g44010	WRKY28, expressed	Enhanced disease resistance	(Peng et al., 2010)
LOC_Os07g40290	OsGH3.8 - Probable indole-3-acetic acid-amido synthetase, expressed	enhanced disease resistance	(Ding et al., 2008)
LOC_Os08g06280	LSD1 zinc finger domain containing protein, expressed	increased susceptibility to fungus	(Wang et al., 2005)

LOC_Os09g25070	WRKY62, expressed	Increased bacterial susceptibility	(Peng et al., 2008)
LOC_Os11g03300	NAC domain transcription factor, putative, expressed	increased drought tolerance	(Jeong et al., 2010)
LOC_Os12g16720	cytochrome P450 71A1, putative, expressed	Enhanced fungal resistance*	(Fujiwara et al., 2010)

*- suppression of gene expression by knock out; N/A- Not Applicable

Conclusion

Availability of large volumes of genome scale gene expression data and advanced computational techniques enabled us to dissect the complex nature of stress response and examine in-depth the overlap between abiotic and biotic stress responses. Plethora of novel insights reported in this work revealed the overarching roles of major stress regulatory molecules including phytohormones such as ABA and JA/ET, parent compounds of small metabolites like shikmate, transcription factors like WRKY and MYB, and signaling genes like WAKs which are central to the fine-tuning of stress response pathways. Further, the expression patterns rendered by these genes provided molecular basis to classify different stress conditions with high accuracy. Altogether, a number of findings in this study vastly build on the existing scientific knowledge and paves way forward to the comprehensive understanding of stress response that is crucial for development of a rice variety with broad range stress tolerance.

Future directions

A number of challenges and formidable hurdles remain to be met before the vast body of knowledge accumulated translates into a successful broad spectrum stress tolerant rice variety. These include identification of sensors and signaling pathways specific to stresses, comprehensive understanding of the molecular basis of interplay among stresses, identification of key factors in the connection between stress responses and developmental processes, addressing how local (a)biotic stress signals are processed and transduced to other parts of the plant body, and examining long-term plant responses under multiple abiotic stress conditions in nature (Hirayama and Shinozaki, 2010).

Various lists of genes identified in this study can be used as a panel to investigate multiple stress responses and the associated molecular mechanisms in different stress environments. The top regulatory and signaling genes in these lists represent potential candidates to improve multiple stress response as compared to those identified in individual stress studies due to the fact that they are involved in multiple stresses, show very high correlation of co-expression and were able to discriminate between different stress conditions. Studies that knock-out or overexpress one or more of these genes in various combinations of stresses and analysis of the resulting phenotypes using the latest high throughput technologies to examine all of the major molecular layers including epigenome, transcriptome, proteome and metabolome would unravel the significance of these genes in stress response network and demonstrate their utility in stress response engineering. A number of genes with high connectivity, conserved expression but with poor annotation were also identified. Experimental studies elucidating their functional roles would reveal novel stress response mechanisms and provide additional targets with great potential in development of transgenic crops with the desired capabilities. Further, mechanistic insights gained in rice on stress responses would provide anchor points to explore specific stress signaling pathways and orthologous genes in other cereal crops.

Supplementary files

Supplementary figures are provided in the accompanying PowerPoint file named supp_figs.pptx

Supplementary tables are provided in the accompanying Excel file named supp_tables.xlsx

References

Adie BA, Perez-Perez J, Perez-Perez MM, Godoy M, Sanchez-Serrano JJ, Schmelz EA et al. (2007) ABA is an essential signal for plant resistance to pathogens affecting JA biosynthesis and the activation of defenses in Arabidopsis. *The Plant cell* **19**, 1665-81.

Ahuja I, de Vos RC, Bones AM, and Hall RD (2010) Plant molecular stress responses face climate change. *Trends in plant science* **15**, 664-74.

Akashi K, Miyake C, and Yokota A (2001) Citrulline, a novel compatible solute in drought-tolerant wild watermelon leaves, is an efficient hydroxyl radical scavenger. *Febs Letters* **508**, 438-42.

Albert R (2007) Network inference, analysis, and modeling in systems biology. *The Plant cell* **19**, 3327-38.

Allen JD, Xie Y, Chen M, Girard L, and Xiao G (2012) Comparing statistical methods for constructing large scale gene networks. *Plos One* **7**, e29348.

Allen KN and Dunaway-Mariano D (2009) Markers of fitness in a successful enzyme superfamily. *Current Opinion in Structural Biology* **19**, 658-65.

An SH, Sohn KH, Choi HW, Hwang IS, Lee SC, and Hwang BK (2008) Pepper pectin methylesterase inhibitor protein CaPMEI1 is required for antifungal activity, basal disease resistance and abiotic stress tolerance. *Planta* **228**, 61-78.

Antao CM and Malcata FX (2005) Plant serine proteases: biochemical, physiological and molecular features. *Plant physiology and biochemistry : PPB / Societe francaise de physiologie vegetale* **43**, 637-50.

Atkinson NJ and Urwin PE (2012) The interaction of plant biotic and abiotic stresses: from genes to the field. *Journal of experimental botany* **63**, 3523-43.

Bari R and Jones JD (2009) Role of plant hormones in plant defence responses. *Plant molecular biology* **69**, 473-88.

Barrett T and Edgar R (2006) Gene expression omnibus: Microarray data storage, submission, retrieval, and analysis. *DNA Microarrays, Part B: Databases and Statistics* **411**, 352-69.

Barrett T, Troup DB, Wilhite SE, Ledoux P, Rudnev D, Evangelista C et al. (2007) NCBI GEO: mining tens of millions of expression profiles--database and tools update. *Nucleic acids research* **35**, D760-5.

Bolouri-Moghaddam MR, Le Roy K, Xiang L, Rolland F, and Van den Ende W (2010) Sugar signalling and antioxidant network connections in plant cells. *The FEBS journal* **277**, 2022-37.

Boyko A, Blevins T, Yao Y, Golubov A, Bilichak A, Ilnytskyy Y et al. (2010) Transgenerational adaptation of Arabidopsis to stress requires DNA methylation and the function of Dicer-like proteins. *Plos One* **5**, e9514.

Breitling R, Armengaud P, Amtmann A, and Herzyk P (2004) Rank products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments. *Febs Letters* **573**, 83-92.

Brossa R, Lopez-Carbonell M, Jubany-Mari T, and Alegre L (2011) Interplay Between Abscisic Acid and Jasmonic Acid and its Role in Water-oxidative Stress in Wild-type, ABA-deficient, JA-deficient, and Ascorbate-deficient Arabidopsis Plants. *Journal of Plant Growth Regulation* **30**, 322-33.

Cao FY, Yoshioka K, and Desveaux D (2011) The roles of ABA in plant-pathogen interactions. *Journal of plant research* **124**, 489-99.

Capell T, Bassie L, and Christou P (2004) Modulation of the polyamine biosynthetic pathway in transgenic rice confers tolerance to drought stress. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 9909-14.

Chae K, Isaacs CG, Reeves PH, Maloney GS, Muday GK, Nagpal P et al. (2012) Arabidopsis SMALL AUXIN UP RNA63 promotes hypocotyl and stamen filament elongation. *The Plant journal : for cell and molecular biology* **71**, 684-97.
Chaves MM, Flexas J, and Pinheiro C (2009) Photosynthesis under drought and salt stress: regulation mechanisms from whole plant to cell. *Annals of Botany* **103**, 551-60.
Chen Z, Hong X, Zhang H, Wang Y, Li X, Zhu JK et al. (2005) Disruption of the cellulose synthase gene, AtCesA8/IRX1, enhances drought and osmotic stress tolerance in Arabidopsis. *The Plant journal : for cell and molecular biology* **43**, 273-83.
Cheng Y, Zhou Y, Yang Y, Chi YJ, Zhou J, Chen JY et al. (2012) Structural and functional analysis of VQ motif-containing proteins in Arabidopsis as interacting proteins of WRKY transcription factors. *Plant Physiology* **159**, 810-25.
Cheng Z, Sun L, Qi T, Zhang B, Peng W, Liu Y et al. (2011) The bHLH transcription factor MYC3 interacts with the Jasmonate ZIM-domain proteins to mediate jasmonate response in Arabidopsis. *Molecular Plant* **4**, 279-88.
Childs KL, Davidson RM, and Buell CR (2011) Gene Coexpression Network Analysis as a Source of Functional Annotation for Rice Genes. *Plos One* **6**,
Chinnusamy V, Schumaker K, and Zhu JK (2004) Molecular genetic perspectives on cross-talk and specificity in abiotic stress signalling in plants. *Journal of experimental botany* **55**, 225-36.
Choudhary MK, Basu D, Datta A, Chakraborty N, and Chakraborty S (2009) Dehydration-responsive nuclear proteome of rice (Oryza sativa L.) illustrates protein network, novel regulators of cellular adaptation, and evolutionary perspective. *Molecular & cellular proteomics : MCP* **8**, 1579-98.
Chung HS and Howe GA (2009) A critical role for the TIFY motif in repression of jasmonate signaling by a stabilized splice variant of the JASMONATE ZIM-domain protein JAZ10 in Arabidopsis. *The Plant cell* **21**, 131-45.
Chung HS, Koo AJ, Gao X, Jayanty S, Thines B, Jones AD et al. (2008) Regulation and function of Arabidopsis JASMONATE ZIM-domain genes in response to wounding and herbivory. *Plant Physiology* **146**, 952-64.
Cohen D, Bogeat-Triboulot MB, Tisserant E, Balzergue S, Martin-Magniette ML, Lelandais G et al. (2010) Comparative transcriptomics of drought responses in Populus: a meta-analysis of genome-wide expression profiling in mature leaves and root apices across two genotypes. *Bmc Genomics* **11**, 630.
Cooper B, Clarke JD, Budworth P, Kreps J, Hutchison D, Park S et al. (2003) A network of rice genes associated with stress response and seed development. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 4945-50.
Cui X and Luan S (2012) A new wave of hormone research: crosstalk mechanisms. *Molecular Plant* **5**, 959-60.
Dai X, Xu Y, Ma Q, Xu W, Wang T, Xue Y et al. (2007) Overexpression of an R1R2R3 MYB gene, OsMYB3R-2, increases tolerance to freezing, drought, and salt stress in transgenic Arabidopsis. *Plant Physiology* **143**, 1739-51.
Dai X and Zhao PX (2011) psRNATarget: a plant small RNA target analysis server. *Nucleic acids research* **39**, W155-9.
Dat J, Vandenabeele S, Vranova E, Van Montagu M, Inze D, and Van Breusegem F (2000) Dual action of the active oxygen species during plant stress responses. *Cellular and molecular life sciences : CMLS* **57**, 779-95.

Demmig-Adams B, Cohu CM, Amiard V, van Zadelhoff G, Veldink GA, Muller O et al. (2013) Emerging trade-offs – impact of photoprotectants (PsbS, xanthophylls, and vitamin E) on oxylipins as regulators of development and defense. *New Phytologist* **197: 720–729**,
Diaz-Uriarte R and Alvarez de Andres S (2006) Gene selection and classification of microarray data using random forest. *BMC bioinformatics* **7**, 3.
Dietrich RA, Richberg MH, Schmidt R, Dean C, and Dang JL (1997) A Novel Zinc Finger Protein Is Encoded by the Arabidopsis LSD1 Gene and Functions as a Negative Regulator of Plant Cell Death. *Cell* **88**, 685-94.
Ding X, Cao Y, Huang L, Zhao J, Xu C, Li X et al. (2008) Activation of the indole-3-acetic acid-amido synthetase GH3-8 suppresses expansin expression and promotes salicylate- and jasmonate-independent basal immunity in rice. *The Plant cell* **20**, 228-40.
Downs GS, Bi YM, Colasanti J, Wu W, Chen X, Zhu T et al. (2013) A developmental transcriptional network for Zea mays defines coexpression modules. *Plant Physiology*
Droillard M, Boudsocq M, Barbier-Brygoo H, and Lauriere C (2002) Different protein kinase families are activated by osmotic stresses in Arabidopsis thaliana cell suspensions. Involvement of the MAP kinases AtMPK3 and AtMPK6. *Febs Letters* **527**, 43-50.
Du Z, Zhou X, Ling Y, Zhang Z, and Su Z (2010) agriGO: a GO analysis toolkit for the agricultural community. *Nucleic acids research* **38**, W64-70.
Duan L, Dietrich D, Ng CH, Chan PM, Bhalerao R, Bennett MJ et al. (2013) Endodermal ABA Signaling Promotes Lateral Root Quiescence during Salt Stress in Arabidopsis Seedlings. *The Plant cell*
Epple P, Mack AA, Morris VR, and Dangl JL (2003) Antagonistic control of oxidative stress-induced cell death in Arabidopsis by two related, plant-specific zinc finger proteins. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 6831-6.
Fattash I, Voss B, Reski R, Hess WR, and Frank W (2007) Evidence for the rapid expansion of microRNA-mediated regulation in early land plant evolution. *Bmc Plant Biology* **7**, 13.
Fernandes H, Michalska K, Sikorski M, and Jaskolski M (2013) Structural and functional aspects of PR-10 proteins. *The FEBS journal* **280**, 1169-99.
Feuillet C and Keller B (2002) Comparative genomics in the grass family: molecular characterization of grass genome structure and evolution. *Annals of Botany* **89**, 3-10.
Ficklin SP and Feltus FA (2011) Gene coexpression network alignment and conservation of gene modules between two grass species: maize and rice. *Plant Physiology* **156**, 1244-56.
Ficklin SP, Luo F, and Feltus FA (2010) The association of multiple interacting genes with specific phenotypes in rice using gene coexpression networks. *Plant Physiology* **154**, 13-24.
Finka A, Mattoo RU, and Goloubinoff P (2011) Meta-analysis of heat- and chemically upregulated chaperone genes in plant and human cells. *Cell stress & chaperones* **16**, 15-31.
Foley JA, Ramankutty N, Brauman KA, Cassidy ES, Gerber JS, Johnston M et al. (2011) Solutions for a cultivated planet. *Nature* **478**, 337-42.
Frank E, Hall M, Trigg L, Holmes G, and Witten IH (2004) Data mining in bioinformatics using Weka. *Bioinformatics* **20**, 2479-81.
Frye CA, Tang DZ, and Innes RW (2001) Negative regulation of defense responses in plants by a conserved MAPKK kinase. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 373-8.
Fu J, Liu H, Li Y, Yu H, Li X, Xiao J et al. (2011) Manipulating broad-spectrum disease resistance by suppressing pathogen-induced auxin accumulation in rice. *Plant Physiology* **155**, 589-602.

Fujita M, Fujita Y, Noutoshi Y, Takahashi F, Narusaka Y, Yamaguchi-Shinozaki K et al. (2006) Crosstalk between abiotic and biotic stress responses: a current view from the points of convergence in the stress signaling networks. *Current opinion in plant biology* **9**, 436-42.
Fujiwara T, Maisonneuve S, Isshiki M, Mizutani M, Chen L, Wong HL et al. (2010) Sekiguchi lesion gene encodes a cytochrome P450 monooxygenase that catalyzes conversion of tryptamine to serotonin in rice. *The Journal of biological chemistry* **285**, 11308-13.
Fukao T and Xiong L (2013) Genetic mechanisms conferring adaptation to submergence and drought in rice: simple or complex? *Current opinion in plant biology*
Furey TS, Cristianini N, Duffy N, Bednarski DW, Schummer M, and Haussler D (2000) Support vector machine classification and validation of cancer tissue samples using microarray expression data. *Bioinformatics* **16**, 906-14.
Gadjev I, Vanderauwera S, Gechev TS, Laloi C, Minkov IN, Shulaev V et al. (2006) Transcriptomic footprints disclose specificity of reactive oxygen species signaling in Arabidopsis. *Plant Physiology* **141**, 436-45.
Garg AK, Kim JK, Owens TG, Ranwala AP, Choi YD, Kochian LV et al. (2002) Trehalose accumulation in rice plants confers high tolerance levels to different abiotic stresses. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 15898-903.
Gautier L, Cope L, Bolstad BM, and Irizarry RA (2004) affy--analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* **20**, 307-15.
Gendler K, Paulsen T, and Napoli C (2008) ChromDB: the chromatin database. *Nucleic acids research* **36**, D298-302.
Gfeller A, Dubugnon L, Liechti R, and Farmer EE (2010) Jasmonate Biochemical Pathway. *Science Signaling* **3**,
Ghanekar R, Srinivasasainagendra V, and Page GP (2008) Cross-chip probe matching tool: A web-based tool for linking microarray probes within and across plant species. *International journal of plant genomics* **2008**, 451327.
Godfray HC (2011) Food for thought. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 19845-6.
Godfray HCJ, Beddington JR, Crute IR, Haddad L, Lawrence D, Muir JF et al. (2010) Food Security: The Challenge of Feeding 9 Billion People. *Science* **327**, 812-8.
Goff SA, Ricke D, Lan TH, Presting G, Wang R, Dunn M et al. (2002) A draft sequence of the rice genome (Oryza sativa L. ssp. japonica). *Science* **296**, 92-100.
Gorovits R, Akad F, Beery H, Vidavsky F, Mahadav A, and Czosnek H (2007) Expression of stress-response proteins upon whitefly-mediated inoculation of Tomato yellow leaf curl virus in susceptible and resistant tomato plants. *Molecular Plant-Microbe Interactions* **20**, 1376-83.
Grennan AK (2006) Genevestigator. Facilitating Web-based gene-expression analysis. *Plant Physiology* **141**, 1164-6.
Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, and Enright AJ (2006) miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic acids research* **34**, D140-4.
Griffiths S, Dunford RP, Coupland G, and Laurie DA (2003) The evolution of CONSTANS-like gene families in barley, rice, and Arabidopsis. *Plant Physiology* **131**, 1855-67.
Hadiarto T and Tran LSP (2011) Progress studies of drought-responsive genes in rice. *Plant Cell Reports* **30**, 297-310.
Hamada K, Hongo K, Suwabe K, Shimizu A, Nagayama T, Abe R et al. (2011) OryzaExpress: an integrated database of gene expression networks and omics annotations in rice. *Plant & cell physiology* **52**, 220-9.

Hand SC, Menze MA, Toner M, Boswell L, and Moore D (2011) LEA Proteins During Water Stress: Not Just for Plants Anymore. *Annual Review of Physiology, Vol 73* **73**, 115-34.
Harb A, Krishnan A, Ambavaram MM, and Pereira A (2010) Molecular and physiological analysis of drought stress in Arabidopsis reveals early responses leading to acclimation in plant growth. *Plant Physiology* **154**, 1254-71.
Hattori Y, Nagai K, Furukawa S, Song XJ, Kawano R, Sakakibara H et al. (2009) The ethylene response factors SNORKEL1 and SNORKEL2 allow rice to adapt to deep water. *Nature* **460**, 1026-30.
Hayden CA and Jorgensen RA (2007) Identification of novel conserved peptide uORF homology groups in Arabidopsis and rice reveals ancient eukaryotic origin of select groups and preferential association with transcription factor-encoding genes. *BMC biology* **5**, 32.
He G, Elling AA, and Deng XW (2011) The epigenome and plant development. *Annual review of plant biology* **62**, 411-35.
Hedenfalk I, Duggan D, Chen Y, Radmacher M, Bittner M, Simon R et al. (2001) Gene-expression profiles in hereditary breast cancer. *The New England journal of medicine* **344**, 539-48.
Hernandez-Blanco C, Feng DX, Hu J, Sanchez-Vallet A, Deslandes L, Llorente F et al. (2007) Impairment of cellulose synthases required for Arabidopsis secondary cell wall formation enhances disease resistance. *The Plant cell* **19**, 890-903.
Hirayama T and Shinozaki K (2010) Research on plant abiotic stress responses in the post-genome era: past, present and future. *The Plant journal : for cell and molecular biology* **61**, 1041-52.
Hong F and Breitling R (2008) A comparison of meta-analysis methods for detecting differentially expressed genes in microarray experiments. *Bioinformatics* **24**, 374-82.
Hong F, Breitling R, McEntee CW, Wittner BS, Nemhauser JL, and Chory J (2006) RankProd: a bioconductor package for detecting differentially expressed genes in meta-analysis. *Bioinformatics* **22**, 2825-7.
Hossain MA, Cho JI, Han M, Ahn CH, Jeon JS, An G et al. (2010) The ABRE-binding bZIP transcription factor OsABF2 is a positive regulator of abiotic stress and ABA signaling in rice. *Journal of plant physiology* **167**, 1512-20.
Hsueh HM, Zhou DW, and Tsai CA (2013) Random forests-based differential analysis of gene sets for gene expression data. *Gene* **518**, 179-86.
Hu H, Dai M, Yao J, Xiao B, Li X, Zhang Q et al. (2006) Overexpressing a NAM, ATAF, and CUC (NAC) transcription factor enhances drought resistance and salt tolerance in rice. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 12987-92.
Hu H, You J, Fang Y, Zhu X, Qi Z, and Xiong L (2008) Characterization of transcription factor gene SNAC2 conferring cold and salt tolerance in rice. *Plant molecular biology* **67**, 169-81.
Huang DW, Sherman BT, and Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols* **4**, 44-57.
Huang M and Guo Z (2005) Responses of antioxidative system to chilling stress in two rice cultivars differing in sensitivity. *Biologia Plantarum* **49**, 81-4.
Hwang I and Sheen J (2001) Two-component circuitry in Arabidopsis cytokinin signal transduction. *Nature* **413**, 383-9.
Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B, and Speed TP (2003) Summaries of Affymetrix GeneChip probe level data. *Nucleic acids research* **31**, e15.
IRRI WRS In *http://www.irri.org/science/ricestat/index.asp*. (ed.), Vol. pp.

Islam MA, Du H, Ning J, Ye H, and Xiong L (2009) Characterization of Glossy1-homologous genes in rice involved in leaf wax accumulation and drought resistance. *Plant molecular biology* **70**, 443-56.
Ito Y, Katsura K, Maruyama K, Taji T, Kobayashi M, Seki M et al. (2006) Functional analysis of rice DREB1/CBF-type transcription factors involved in cold-responsive gene expression in transgenic rice. *Plant & cell physiology* **47**, 141-53.
Jaiswal P, Ni J, Yap I, Ware D, Spooner W, Youens-Clark K et al. (2006) Gramene: a bird's eye view of cereal genomes. *Nucleic acids research* **34**, D717-23.
Jan A, Maruyama K, Todaka D, Kidokoro S, Abo M, Yoshimura E et al. (2013) OsTZF1, a CCCH-tandem zinc finger protein, confers delayed senescence and stress tolerance in rice by regulating stress-related genes. *Plant Physiology* **161**, 1202-16.
Jarvis P (2008) Targeting of nucleus-encoded proteins to chloroplasts in plants. *The New phytologist* **179**, 257-85.
Jeong JS, Kim YS, Baek KH, Jung H, Ha SH, Do Choi Y et al. (2010) Root-specific expression of OsNAC10 improves drought tolerance and grain yield in rice under field drought conditions. *Plant Physiology* **153**, 185-97.
Jiang Z, Liu X, Peng Z, Wan Y, Ji Y, He W et al. (2011) AHD2.0: an update version of Arabidopsis Hormone Database for plant systematic studies. *Nucleic acids research* **39**, D1123-9.
Jirschitzka J, Mattern DJ, Gershenzon J, and D'Auria JC (2013) Learning from nature: new approaches to the metabolic engineering of plant defense pathways. *Current opinion in biotechnology* **24**, 320-8.
Johnson KL, Jones BJ, Bacic A, and Schultz CJ (2003) The fasciclin-like arabinogalactan proteins of Arabidopsis. A multigene family of putative cell adhesion molecules. *Plant Physiology* **133**, 1911-25.
Jorgensen RA and Dorantes-Acosta AE (2012) Conserved Peptide Upstream Open Reading Frames are Associated with Regulatory Genes in Angiosperms. *Frontiers in plant science* **3**, 191.
Jung KH, An G, and Ronald PC (2008) Towards a better bowl of rice: assigning function to tens of thousands of rice genes. *Nature reviews. Genetics* **9**, 91-101.
Kapushesky M, Adamusiak T, Burdett T, Culhane A, Farne A, Filippov A et al. (2012) Gene Expression Atlas update-a value-added database of microarray and sequencing-based functional genomics experiments. *Nucleic acids research* **40**, D1077-D81.
Kauffmann A, Gentleman R, and Huber W (2009) arrayQualityMetrics--a bioconductor package for quality assessment of microarray data. *Bioinformatics* **25**, 415-6.
Kawasaki S, Miyake C, Kohchi T, Fujii S, Uchida M, and Yokota A (2000) Responses of wild watermelon to drought stress: accumulation of an ArgE homologue and citrulline in leaves during water deficits. *Plant & cell physiology* **41**, 864-73.
Kim DH, Yamaguchi S, Lim S, Oh E, Park J, Hanada A et al. (2008) SOMNUS, a CCCH-type zinc finger protein in Arabidopsis, negatively regulates light-dependent seed germination downstream of PIL5. *The Plant cell* **20**, 1260-77.
Kim TH, Hauser F, Ha T, Xue S, Bohmer M, Nishimura N et al. (2011) Chemical genetics reveals negative regulation of abscisic acid signaling by a plant immune response pathway. *Current biology : CB* **21**, 990-7.
Klose RJ, Kallin EM, and Zhang Y (2006) JmjC-domain-containing proteins and histone demethylation. *Nature reviews. Genetics* **7**, 715-27.
Koga H, Dohi K, and Mori M (2004) Abscisic acid and low temperatures suppress the whole plant-specific resistance reaction of rice plants to the infection of Magnaporthe grisea. *Physiological and Molecular Plant Pathology* **65**, 3-9.

Kohorn BD and Kohorn SL (2012) The cell wall-associated kinases, WAKs, as pectin receptors. *Frontiers in plant science* **3**, 88.
Kozomara A and Griffiths-Jones S (2011) miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic acids research* **39**, D152-7.
Kumar B, Singla-Pareek SL, and Sopory SK (2010) Glutathione Homeostasis: Crucial for Abiotic Stress Tolerance in Plants. *Abiotic Stress Adaptation in Plants* 263-82.
Lai Z, Li Y, Wang F, Cheng Y, Fan B, Yu JQ et al. (2011) Arabidopsis sigma factor binding proteins are activators of the WRKY33 transcription factor in plant defense. *The Plant cell* **23**, 3824-41.
Langfelder P and Horvath S (2008) WGCNA: an R package for weighted correlation network analysis. *BMC bioinformatics* **9**, 559.
Lee I, Seo YS, Coltrane D, Hwang S, Oh T, Marcotte EM et al. (2011) Genetic dissection of the biotic stress response using a genome-scale gene network for rice. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 18548-53.
Lee SC and Luan S (2012) ABA signal transduction at the crossroad of biotic and abiotic stress responses. *Plant, cell & environment* **35**, 53-60.
Lee TH, Kim YK, Pham TT, Song SI, Kim JK, Kang KY et al. (2009) RiceArrayNet: a database for correlating gene expression from transcriptome profiling, and its application to the analysis of coexpressed genes in rice. *Plant Physiology* **151**, 16-33.
Lenka SK, Katiyar A, Chinnusamy V, and Bansal KC (2011) Comparative analysis of drought-responsive transcriptome in Indica rice genotypes with contrasting drought tolerance. *Plant Biotechnology Journal* **9**, 315-27.
Li H, Zhou SY, Zhao WS, Su SC, and Peng YL (2009) A novel wall-associated receptor-like protein kinase gene, OsWAK1, plays important roles in rice blast disease resistance. *Plant molecular biology* **69**, 337-46.
Li Z and Thomas TL (1998) PEI1, an embryo-specific zinc finger protein gene required for heart-stage embryo formation in Arabidopsis. *The Plant cell* **10**, 383-98.
Liang Y, Zhang F, Wang J, Joshi T, Wang Y, and Xu D (2011) Prediction of drought-resistant genes in Arabidopsis thaliana using SVM-RFE. *Plos One* **6**, e21750.
Lichtenthaler HK (1984) Differences in morphology and chemical composition of leaves grown at different light intensities and qualitites. In Control of Leaf Growth. *Cambridge University Press. Cambridge, United Kingdom* **201-222**,
Lichtenthaler HK (1998) The stress concept in plants: An introduction. *Stress of Life* **851**, 187-98.
Lindow M, Jacobsen A, Nygaard S, Mang Y, and Krogh A (2007) Intragenomic matching reveals a huge potential for miRNA-mediated regulation in plants. *PLoS computational biology* **3**, e238.
Liu WY, Wang MM, Huang J, Tang HJ, Lan HX, and Zhang HS (2009) The OsDHODH1 gene is involved in salt and drought tolerance in rice. *Journal of Integrative Plant Biology* **51**, 825-33.
Liu X, Bai X, Wang X, and Chu C (2007) OsWRKY71, a rice transcription factor, is involved in rice defense response. *Journal of plant physiology* **164**, 969-79.
Liu Z, Xie M, Yao Z, Niu Y, Bu Y, and Gao C (2013) Three meta-analyses define a set of commonly overexpressed genes from microarray datasets on astrocytomas. *Molecular neurobiology* **47**, 325-36.
Lloyd J and Meinke D (2012) A comprehensive dataset of genes with a loss-of-function mutant phenotype in Arabidopsis. *Plant Physiology* **158**, 1115-29.
Lodha TD and Basak J (2012) Plant-pathogen interactions: what microarray tells about it? *Molecular biotechnology* **50**, 87-97.
López-Gómez M and Lluch C (2012) Trehalose and Abiotic Stress Tolerance. *Abiotic Stress Responses in Plants* 253-65.

Lukk M, Kapushesky M, Nikkila J, Parkinson H, Goncalves A, Huber W et al. (2010) A global map of human gene expression. *Nature biotechnology* **28**, 322-4.
Lutz W and Samir KC (2010) Dimensions of global population projections: what do we know about future population trends and structures? *Philosophical Transactions of the Royal Society B-Biological Sciences* **365**, 2779-91.
Ma H and Zhao J (2010) Genome-wide identification, classification, and expression analysis of the arabinogalactan protein gene family in rice (Oryza sativa L.). *Journal of experimental botany* **61**, 2647-68.
Maeda H and Dudareva N (2012) The Shikimate Pathway and Aromatic Amino Acid Biosynthesis in Plants. *Annual Review of Plant Biology, Vol 63* **63**, 73-105.
Mang HG, Qian W, Zhu Y, Qian J, Kang HG, Klessig DF et al. (2012) Abscisic acid deficiency antagonizes high-temperature inhibition of disease resistance through enhancing nuclear accumulation of resistance proteins SNC1 and RPS4 in Arabidopsis. *The Plant cell* **24**, 1271-84.
Mao XZ, Cai T, Olyarchuk JG, and Wei LP (2005) Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary. *Bioinformatics* **21**, 3787-93.
Mayer MP and Bukau B (2005) Hsp70 chaperones: cellular functions and molecular mechanism. *Cellular and molecular life sciences : CMLS* **62**, 670-84.
Meier S, Bastian R, Donaldson L, Murray S, Bajic V, and Gehring C (2008) Co-expression and promoter content analyses assign a role in biotic and abiotic stress responses to plant natriuretic peptides. *Bmc Plant Biology* **8**, 24.
Melotto M, Underwood W, Koczan J, Nomura K, and He SY (2006) Plant stomata function in innate immunity against bacterial invasion. *Cell* **126**, 969-80.
Mittler R and Blumwald E (2010) Genetic engineering for modern agriculture: challenges and perspectives. *Annual review of plant biology* **61**, 443-62.
Mizoi J and Yamaguchi-Shinozaki K (2013) Molecular approaches to improve rice abiotic stress tolerance. *Methods in molecular biology* **956**, 269-83.
Mochida K, Uehara-Yamaguchi Y, Yoshida T, Sakurai T, and Shinozaki K (2011) Global landscape of a co-expressed gene network in barley and its application to gene discovery in Triticeae crops. *Plant & cell physiology* **52**, 785-803.
Moons A (2005) Regulatory and functional interactions of plant growth regulators and plant glutathione S-transferases (GSTs). *Vitamins and hormones* **72**, 155-202.
Moons A (2008) Transcriptional profiling of the PDR gene family in rice roots in response to plant growth regulators, redox perturbations and weak organic acid stresses. *Planta* **229**, 53-71.
Mosolov VV and Valueva TA (2011) [Inhibitors of proteolytic enzymes under abiotic stresses in plants (review)]. *Prikladnaia biokhimiia i mikrobiologiia* **47**, 501-7.
Muench DG, Zhang C, and Dahodwala M (2012) Control of cytoplasmic translation in plants. *Wiley interdisciplinary reviews. RNA* **3**, 178-94.
Nagamura Y, Antonio BA, Sato Y, Miyao A, Namiki N, Yonemaru J et al. (2011) Rice TOGO Browser: A platform to retrieve integrated information on rice functional and applied genomics. *Plant & cell physiology* **52**, 230-7.
Naika M, Shameer K, Mathew OK, Gowda R, and Sowdhamini R (2013) STIFDB2: an updated version of plant stress-responsive transcription factor database with additional stress signals, stress-responsive transcription factor binding sites and stress-responsive genes in Arabidopsis and rice. *Plant & cell physiology* **54**, e8.

Nakashima K, Tran LS, Van Nguyen D, Fujita M, Maruyama K, Todaka D et al. (2007) Functional analysis of a NAC-type transcription factor OsNAC6 involved in abiotic and biotic stress-responsive gene expression in rice. *The Plant journal : for cell and molecular biology* **51**, 617-30.
Narsai R, Wang C, Chen J, Wu J, Shou H, and Whelan J (2013) Antagonistic, overlapping and distinct responses to biotic stress in rice (Oryza sativa) and interactions with abiotic stress. *Bmc Genomics* **14**, 93.
Newton AC, Johnson SN, and Gregory PJ (2011) Implications of climate change for diseases, crop yields and food security. *Euphytica* **179**, 3-18.
Obayashi T, Hayashi S, Saeki M, Ohta H, and Kinoshita K (2009) ATTED-II provides coexpressed gene networks for Arabidopsis. *Nucleic acids research* **37**, D987-91.
Oh SJ, Kim YS, Kwon CW, Park HK, Jeong JS, and Kim JK (2009) Overexpression of the transcription factor AP37 in rice improves grain yield under drought conditions. *Plant Physiology* **150**, 1368-79.
Oliveros JC (2007) An interactive tool for comparing lists with Venn Diagrams. In (ed.), Vol. pp.
Pandey SP and Somssich IE (2009) The role of WRKY transcription factors in plant immunity. *Plant Physiology* **150**, 1648-55.
Pantin F, Monnet F, Jannaud D, Costa JM, Renaud J, Muller B et al. (2013) The dual effect of abscisic acid on stomata. *The New phytologist* **197**, 65-72.
Park JE, Park JY, Kim YS, Staswick PE, Jeon J, Yun J et al. (2007) GH3-mediated auxin homeostasis links growth regulation with stress adaptation response in Arabidopsis. *The Journal of biological chemistry* **282**, 10036-46.
Peng Y, Bartley LE, Canlas P, and Ronald PC (2010) OsWRKY IIa Transcription Factors Modulate Rice Innate Immunity. *Rice* **3**, 36-42.
Peng Y, Bartley LE, Chen X, Dardick C, Chern M, Ruan R et al. (2008) OsWRKY62 is a negative regulator of basal and Xa21-mediated defense against Xanthomonas oryzae pv. oryzae in rice. *Molecular Plant* **1**, 446-58.
Perez-Enciso M and Tenenhaus M (2003) Prediction of clinical outcome with microarray data: a partial least squares discriminant analysis (PLS-DA) approach. *Human genetics* **112**, 581-92.
Perez-Rodriguez P, Riano-Pachon DM, Correa LG, Rensing SA, Kersten B, and Mueller-Roeber B (2010) PlnTFDB: updated content and new features of the plant transcription factor database. *Nucleic acids research* **38**, D822-7.
Pieterse CM, Leon-Reyes A, Van der Ent S, and Van Wees SC (2009) Networking by small-molecule hormones in plant immunity. *Nature chemical biology* **5**, 308-16.
Puhakainen T, Hess MW, Makela P, Svensson J, Heino P, and Palva ET (2004) Overexpression of multiple dehydrin genes enhances tolerance to freezing stress in Arabidopsis. *Plant molecular biology* **54**, 743-53.
Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C et al. (2012) The Pfam protein families database. *Nucleic acids research* **40**, D290-301.
Qu LJ, Chen J, Liu M, Pan N, Okamoto H, Lin Z et al. (2003) Molecular cloning and functional analysis of a novel type of Bowman-Birk inhibitor gene family in rice. *Plant Physiology* **133**, 560-70.
Rabbani MA, Maruyama K, Abe H, Khan MA, Katsura K, Ito Y et al. (2003) Monitoring expression profiles of rice genes under cold, drought, and high-salinity stresses and abscisic acid application using cDNA microarray and RNA gel-blot analyses. *Plant Physiology* **133**, 1755-67.
Rakwal R, Kumar Agrawal G, and Jwa NS (2001) Characterization of a rice (Oryza sativa L.) Bowman-Birk proteinase inhibitor: tightly light regulated induction in response to cut, jasmonic acid, ethylene and protein phosphatase 2A inhibitors. *Gene* **263**, 189-98.

Ray S, Dansana PK, Giri J, Deveshwar P, Arora R, Agarwal P et al. (2011) Modulation of transcription factor and metabolic pathway genes in response to water-deficit stress in rice. *Functional & Integrative Genomics* **11**, 157-78.
Reynolds MP, Hellin J, Govaerts B, Kosina P, Sonder K, Hobbs P et al. (2012) Global crop improvement networks to bridge technology gaps. *Journal of experimental botany* **63**, 1-12.
Ribot C, Hirsch J, Balzergue S, Tharreau D, Notteghem JL, Lebrun MH et al. (2008) Susceptibility of rice to the blast fungus, Magnaporthe grisea. *Journal of plant physiology* **165**, 114-24.
Riechmann JL and Meyerowitz EM (1997) MADS domain proteins in plant development. *Biological chemistry* **378**, 1079-101.
Rock CD, Sakata Y, and Quatrano RS (2010) Stress signaling I: The role of abscisic acid (ABA). *Abiotic Stress Adaptation in Plants* 33-73.
Rodrigo G, Carrera J, Ruiz-Ferrer V, del Toro FJ, Llave C, Voinnet O et al. (2012) A meta-analysis reveals the commonalities and differences in Arabidopsis thaliana response to different viral pathogens. *Plos One* **7**, e40526.
Rouard M, Guignon V, Aluome C, Laporte MA, Droc G, Walde C et al. (2011) GreenPhylDB v2.0: comparative and functional genomics in plants. *Nucleic acids research* **39**, D1095-102.
Rung J and Brazma A (2013) Reuse of public genome-wide gene expression data. *Nature reviews. Genetics* **14**, 89-99.
Rustici G, Kolesnikov N, Brandizi M, Burdett T, Dylag M, Emam I et al. (2013) ArrayExpress update-trends in database growth and links to data analysis tools. *Nucleic acids research* **41**, D987-D90.
Sairam RK, Srivastava GC, Agarwal S, and Meena RC (2005) Differences in antioxidant activity in response to salinity stress in tolerant and susceptible wheat genotypes. *Biologia Plantarum* **49**, 85-91.
Sanchez-Vallet A, Lopez G, Ramos B, Delgado-Cerezo M, Riviere MP, Llorente F et al. (2012) Disruption of abscisic acid signaling constitutively activates Arabidopsis resistance to the necrotrophic fungus Plectosphaerella cucumerina. *Plant Physiology* **160**, 2109-24.
Sasaki K, Iwai T, Hiraga S, Kuroda K, Seo S, Mitsuhara I et al. (2004) Ten rice peroxidases redundantly respond to multiple stresses including infection with rice blast fungus. *Plant & cell physiology* **45**, 1442-52.
Sato Y, Namiki N, Takehisa H, Kamatsuki K, Minami H, Ikawa H et al. (2013) RiceFREND: a platform for retrieving coexpressed gene networks in rice. *Nucleic acids research* **41**, D1214-21.
Seki M, Narusaka M, Ishida J, Nanjo T, Fujita M, Oono Y et al. (2002) Monitoring the expression profiles of 7000 Arabidopsis genes under drought, cold and high-salinity stresses using a full-length cDNA microarray. *The Plant journal : for cell and molecular biology* **31**, 279-92.
Seo YS, Chern M, Bartley LE, Han M, Jung KH, Lee I et al. (2011) Towards establishment of a rice stress response interactome. *PLoS genetics* **7**, e1002020.
Shah J (2003) The salicylic acid loop in plant defense. *Current opinion in plant biology* **6**, 365-71.
Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D et al. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research* **13**, 2498-504.
Sharma R, De Vleesschauwer D, Sharma MK, and Ronald PC (2013) Recent advances in dissecting stress-regulatory crosstalk in rice. *Molecular Plant* **6**, 250-60.
Sharma R, Mohan Singh RK, Malik G, Deveshwar P, Tyagi AK, Kapoor S et al. (2009) Rice cytosine DNA methyltransferases - gene expression profiling during reproductive development and abiotic stress. *The FEBS journal* **276**, 6301-11.

Shen J, Xie K, and Xiong L (2010) Global expression profiling of rice microRNAs by one-tube stem-loop reverse transcription quantitative PCR revealed important roles of microRNAs in abiotic stress responses. *Molecular genetics and genomics : MGG* **284**, 477-88.
Shinozaki K and Yamaguchi-Shinozaki K (2007) Gene networks involved in drought stress response and tolerance. *Journal of experimental botany* **58**, 221-7.
Shu L, Lou Q, Ma C, Ding W, Zhou J, Wu J et al. (2011a) Genetic, proteomic and metabolic analysis of the regulation of energy storage in rice seedlings in response to drought. *Proteomics* **11**, 4122-38.
Shu LB, Lou QJ, Ma CF, Ding W, Zhou J, Wu JH et al. (2011b) Genetic, proteomic and metabolic analysis of the regulation of energy storage in rice seedlings in response to drought. *Proteomics* **11**, 4122-38.
Simon SA and Meyers BC (2011) Small RNA-mediated epigenetic modifications in plants. *Current opinion in plant biology* **14**, 148-55.
Smith BD (1995) In *The emergence of agriculture*. Vol. pp. Scientific American Library New York,
Smith P, Gregory PJ, van Vuuren D, Obersteiner M, Havlik P, Rounsevell M et al. (2010) Competition for land. *Philosophical Transactions of the Royal Society B-Biological Sciences* **365**, 2941-57.
Song XQ, Liu LF, Jiang YJ, Zhang BC, Gao YP, Liu XL et al. (2013) Disruption of Secondary Wall Cellulose Biosynthesis Alters Cadmium Translocation and Tolerance in Rice Plants. *Molecular Plant*
Srinivasasainagendra V, Page GP, Mehta T, Coulibaly I, and Loraine AE (2008) CressExpress: a tool for large-scale mining of expression data from Arabidopsis. *Plant Physiology* **147**, 1004-16.
Stammers DK, Ren J, Leslie K, Nichols CE, Lamb HK, Cocklin S et al. (2001) The structure of the negative transcriptional regulator NmrA reveals a structural superfamily which includes the short-chain dehydrogenase/reductases. *The EMBO journal* **20**, 6619-26.
Staswick PE (2008) JAZing up jasmonate signaling. *Trends in plant science* **13**, 66-71.
Student S and Fujarewicz K (2012) Stable feature selection and classification algorithms for multiclass microarray data. *Biology direct* **7**, 33.
Sun J, Jiang H, Xu Y, Li H, Wu X, Xie Q et al. (2007) The CCCH-type zinc finger proteins AtSZF1 and AtSZF2 regulate salt stress responses in Arabidopsis. *Plant & cell physiology* **48**, 1148-58.
Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Minguez P et al. (2011) The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic acids research* **39**, D561-8.
Takeda S and Matsuoka M (2008) Genetic approaches to crop improvement: responding to environmental and population changes. *Nature Reviews Genetics* **9**, 444-57.
Tao Z, Liu H, Qiu D, Zhou Y, Li X, Xu C et al. (2009) A pair of allelic WRKY genes play opposite roles in rice-bacteria interactions. *Plant Physiology* **151**, 936-48.
Tarutani Y, Shiba H, Iwano M, Kakizaki T, Suzuki G, Watanabe M et al. (2010) Trans-acting small RNA determines dominance relationships in Brassica self-incompatibility. *Nature* **466**, 983-6.
Tester M and Langridge P (2010) Breeding technologies to increase crop production in a changing world. *Science* **327**, 818-22.
Thaler JS, Humphrey PT, and Whiteman NK (2012) Evolution of jasmonate and salicylate signal crosstalk. *Trends in plant science* **17**, 260-70.
To JP, Deruere J, Maxwell BB, Morris VF, Hutchison CE, Ferreira FJ et al. (2007) Cytokinin regulates type-A Arabidopsis Response Regulator activity and protein stability via two-component phosphorelay. *The Plant cell* **19**, 3901-14.

Toda T, Fujii S, Noguchi K, Kazama T, and Toriyama K (2012) Rice MPR25 encodes a pentatricopeptide repeat protein and is essential for RNA editing of nad5 transcripts in mitochondria. *The Plant journal : for cell and molecular biology* **72**, 450-60.
Tohge T, Watanabe M, Hoefgen R, and Fernie AR (2013) The evolution of phenylpropanoid metabolism in the green lineage. *Critical reviews in biochemistry and molecular biology*
Tsai YC, Weir NR, Hill K, Zhang WJ, Kim HJ, Shiu SH et al. (2012) Characterization of Genes Involved in Cytokinin Signaling and Metabolism from Rice. *Plant Physiology* **158**, 1666-84.
Tseng GC, Ghosh D, and Feingold E (2012) Comprehensive literature review and statistical considerations for microarray meta-analysis. *Nucleic acids research*
Tzin V and Galili G (2010) New Insights into the Shikimate and Aromatic Amino Acids Biosynthesis Pathways in Plants. *Molecular Plant* **3**, 956-72.
Upchurch RG (2008) Fatty acid unsaturation, mobilization, and regulation in the response of plants to stress. *Biotechnology letters* **30**, 967-77.
Van Bel M, Proost S, Wischnitzki E, Movahedi S, Scheerlinck C, Van de Peer Y et al. (2012) Dissecting Plant Genomes with the PLAZA Comparative Genomics Platform. *Plant Physiology* **158**, 590-600.
van den Berg RA, Hoefsloot HC, Westerhuis JA, Smilde AK, and van der Werf MJ (2006) Centering, scaling, and transformations: improving the biological information content of metabolomics data. *Bmc Genomics* **7**, 142.
van der Hoorn RA (2008) Plant proteases: from phenotypes to molecular mechanisms. *Annual review of plant biology* **59**, 191-223.
Vinebrooke RD, Cottingham KL, Norberg J, Scheffer M, Dodson SI, Maberly SC et al. (2004) Impacts of multiple stressors on biodiversity and ecosystem functioning: the role of species co-tolerance. *Oikos* **104**, 451-7.
Vinocur B and Altman A (2005) Recent advances in engineering plant tolerance to abiotic stress: achievements and limitations. *Current opinion in biotechnology* **16**, 123-32.
Vlot AC, Dempsey DA, and Klessig DF (2009) Salicylic Acid, a multifaceted hormone to combat disease. *Annual review of phytopathology* **47**, 177-206.
Vogt T (2010) Phenylpropanoid Biosynthesis. *Molecular Plant* **3**, 2-20.
Voinnet O (2009) Origin, biogenesis, and activity of plant microRNAs. *Cell* **136**, 669-87.
Wang D, Pan Y, Zhao X, Zhu L, Fu B, and Li Z (2011a) Genome-wide temporal-spatial gene expression profiling of drought responsiveness in rice. *Bmc Genomics* **12**, 149.
Wang HG, Zhang HL, Gao FH, Li JX, and Li ZC (2007) Comparison of gene expression between upland and lowland rice cultivars under water stress using cDNA microarray. *Theoretical and Applied Genetics* **115**, 1109-26.
Wang L, Pei Z, Tian Y, and He C (2005) OsLSD1, a rice zinc finger protein, regulates programmed cell death and callus differentiation. *Molecular plant-microbe interactions : MPMI* **18**, 375-84.
Wang WS, Pan YJ, Zhao XQ, Dwivedi D, Zhu LH, Ali J et al. (2011b) Drought-induced site-specific DNA methylation and its association with drought tolerance in rice (Oryza sativa L.). *Journal of experimental botany* **62**, 1951-60.
WASDE (2012) U.S. Drought 2012: Farm and Food Impacts http://www.ers.usda.gov/topics/in-the-news/us-drought-2012-farm-and-food-impacts.aspx.
Waters MT, Wang P, Korkaric M, Capper RG, Saunders NJ, and Langdale JA (2009) GLK transcription factors coordinate expression of the photosynthetic apparatus in Arabidopsis. *The Plant cell* **21**, 1109-28.

Weston DJ, Karve AA, Gunter LE, Jawdy SS, Yang X, Allen SM et al. (2011) Comparative physiology and transcriptional networks underlying the heat shock response in Populus trichocarpa, Arabidopsis thaliana and Glycine max. *Plant, cell & environment* **34**, 1488-506.
Wilkinson S, Kudoyarova GR, Veselov DS, Arkhipova TN, and Davies WJ (2012) Plant hormone interactions: innovative targets for crop breeding and management. *Journal of experimental botany* **63**, 3499-509.
Xia J, Mandal R, Sinelnikov IV, Broadhurst D, and Wishart DS (2012) MetaboAnalyst 2.0--a comprehensive server for metabolomic data analysis. *Nucleic acids research* **40**, W127-33.
Xiang Y, Tang N, Du H, Ye H, and Xiong L (2008) Characterization of OsbZIP23 as a key player of the basic leucine zipper transcription factor family for conferring abscisic acid sensitivity and salinity and drought tolerance in rice. *Plant Physiology* **148**, 1938-52.
Xiao B, Huang Y, Tang N, and Xiong L (2007) Over-expression of a LEA gene in rice improves drought resistance under the field conditions. *TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik* **115**, 35-46.
Xiong L and Yang Y (2003) Disease resistance and abiotic stress tolerance in rice are inversely modulated by an abscisic acid-inducible mitogen-activated protein kinase. *The Plant cell* **15**, 745-59.
Xu C and He C (2007) The rice OsLOL2 gene encodes a zinc finger protein involved in rice growth and disease resistance. *Molecular genetics and genomics : MGG* **278**, 85-94.
Xu K, Xu X, Fukao T, Canlas P, Maghirang-Rodriguez R, Heuer S et al. (2006) Sub1A is an ethylene-response-factor-like gene that confers submergence tolerance to rice. *Nature* **442**, 705-8.
Xu ZS, Chen M, Li LC, and Ma YZ (2011) Functions and application of the AP2/ERF transcription factor family in crop improvement. *Journal of Integrative Plant Biology* **53**, 570-85.
Yan H, Kikuchi S, Neumann P, Zhang W, Wu Y, Chen F et al. (2010) Genome-wide mapping of cytosine methylation revealed dynamic DNA methylation patterns associated with genes and centromeres in rice. *The Plant journal : for cell and molecular biology*
Yanhui C, Xiaoyuan Y, Kun H, Meihua L, Jigang L, Zhaofeng G et al. (2006) The MYB transcription factor superfamily of Arabidopsis: expression analysis and phylogenetic comparison with the rice MYB family. *Plant molecular biology* **60**, 107-24.
Ye H, Du H, Tang N, Li X, and Xiong L (2009) Identification and expression profiling analysis of TIFY family genes involved in stress and phytohormone responses in rice. *Plant molecular biology* **71**, 291-305.
Yeung KY and Ruzzo WL (2001) Principal component analysis for clustering gene expression data. *Bioinformatics* **17**, 763-74.
Yim WC, Yu Y, Song K, Jang CS, and Lee BM (2013) PLANEX: the plant co-expression database. *Bmc Plant Biology* **13**, 83.
Yu J, Hu S, Wang J, Wong GK, Li S, Liu B et al. (2002) A draft sequence of the rice genome (Oryza sativa L. ssp. indica). *Science* **296**, 79-92.
Yu S, Ligang C, Liping Z, and Diqiu Y (2010) Overexpression of OsWRKY72 gene interferes in the abscisic acid signal and auxin transport pathway of Arabidopsis. *Journal of Biosciences* **35**, 459-71.
Yuan H and Liu D (2012) Functional disruption of the pentatricopeptide protein SLG1 affects mitochondrial RNA editing, plant development, and responses to abiotic stresses in Arabidopsis. *The Plant journal : for cell and molecular biology* **70**, 432-44.
Zhang J, Li X, He Z, Zhao X, Wang Q, Zhou B et al. (2012a) Molecular character of a phosphatase 2C (PP2C) gene relation to stress tolerance in Arabidopsis thaliana. *Molecular biology reports*

Zhang L, Yu S, Zuo K, Luo L, and Tang K (2012b) Identification of gene modules associated with drought response in rice by network-based analysis. *Plos One* **7**, e33748.
Zhang PJ, Broekgaarden C, Zheng SJ, Snoeren TA, van Loon JJ, Gols R et al. (2013) Jasmonate and ethylene signaling mediate whitefly-induced interference with indirect plant defense in Arabidopsis thaliana. *The New phytologist* **197**, 1291-9.
Zhang S, Chen C, Li L, Meng L, Singh J, Jiang N et al. (2005) Evolutionary expansion, gene structure, and expression of the rice wall-associated kinase gene family. *Plant Physiology* **139**, 1107-24.
Zhang W, Wu Y, Schnable JC, Zeng Z, Freeling M, Crawford GE et al. (2012c) High-resolution mapping of open chromatin in the rice genome. *Genome Research* **22**, 151-62.
Zhang X, Lu X, Shi Q, Xu XQ, Leung HC, Harris LN et al. (2006) Recursive SVM feature selection and sample classification for mass-spectrometry and microarray data. *BMC bioinformatics* **7**, 197.
Zhang Z, Yu J, Li D, Liu F, Zhou X, Wang T et al. (2010) PMRD: plant microRNA database. *Nucleic acids research* **38**, D806-13.
Zhao L, Hu Y, Chong K, and Wang T (2010) ARAG1, an ABA-responsive DREB gene, plays a role in seed germination and drought tolerance of rice. *Annals of Botany* **105**, 401-9.
Zheng X, Chen B, Lu G, and Han B (2009) Overexpression of a NAC transcription factor enhances rice drought and salt tolerance. *Biochemical and Biophysical Research Communications* **379**, 985-9.
Zhou J, Wang X, Jiao Y, Qin Y, Liu X, He K et al. (2007) Global genome expression analysis of rice in response to drought and high-salinity stresses in shoot, flag leaf, and panicle. *Plant molecular biology* **63**, 591-608.
Zhou L, Liu Y, Liu Z, Kong D, Duan M, and Luo L (2010) Genome-wide identification and analysis of drought-responsive microRNAs in Oryza sativa. *Journal of experimental botany* **61**, 4157-68.
Zimmermann P, Hirsch-Hoffmann M, Hennig L, and Gruissem W (2004) GENEVESTIGATOR. Arabidopsis microarray database and analysis toolbox. *Plant Physiology* **136**, 2621-32.